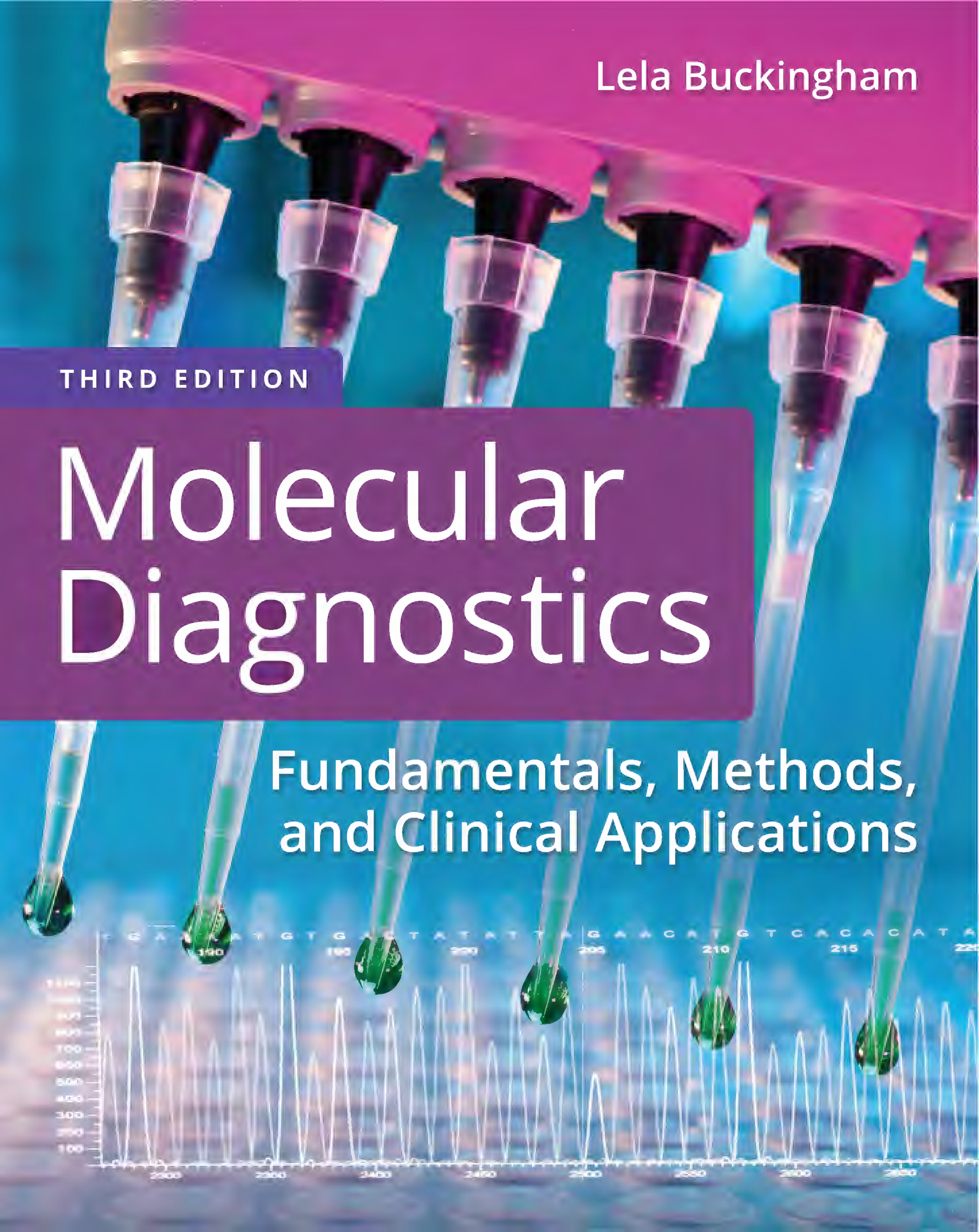


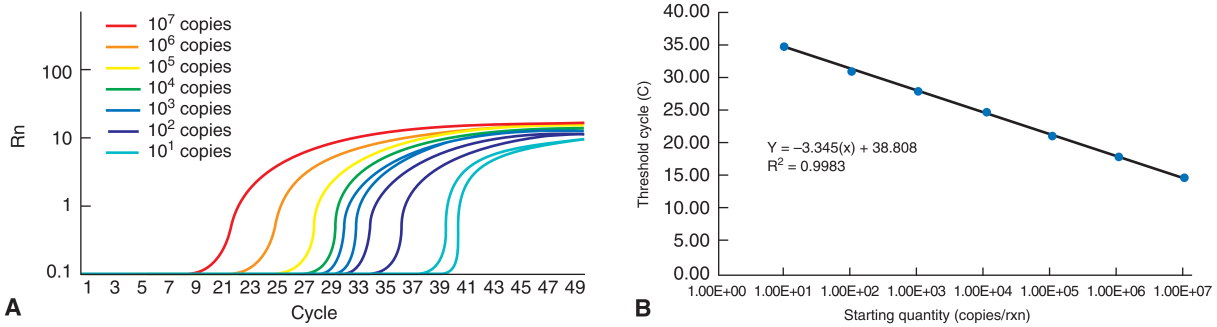
Lela Buckingham

THIRD EDITION

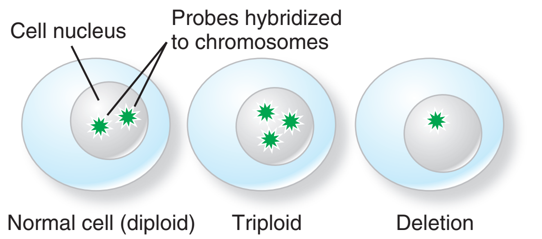
Molecular Diagnostics

Fundamentals, Methods, and Clinical Applications

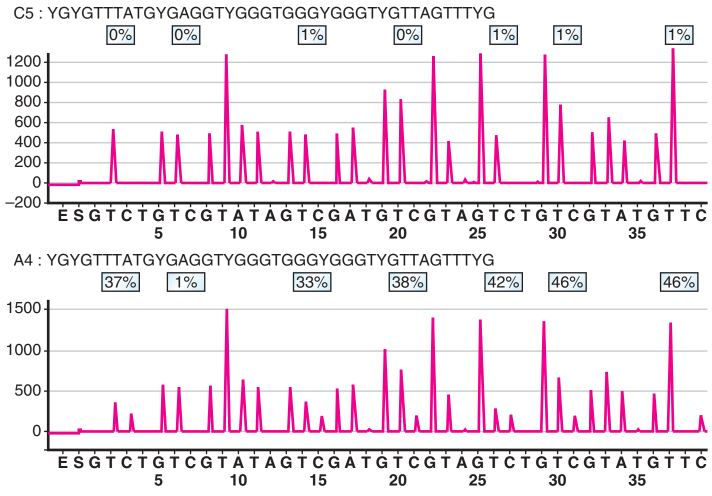
QUICK VISUAL REFERENCE FOR COMMON TECHNIQUES IN MOLECULAR BIOLOGY



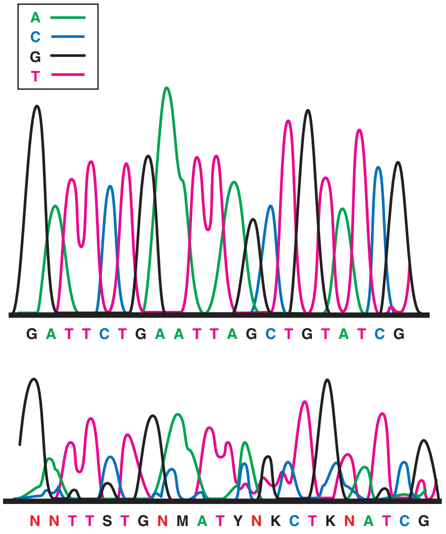
■ Quantitative polymerase chain reaction (PCR)


Cell nucleus
Probes hybridized to chromosomes
Normal cell (diploid)
Triploid
Deletion

■ Fluorescence in situ hybridization (FISH) analysis for normal diploid cell (left), triploidy (center), and deletion (right)



■ DNA methylation at cytosine residues detected by pyrosequencing of bisulfite-treated DNA



■ Examples of good sequence quality (top) and poor sequence quality (bottom)

Continued on inside back cover

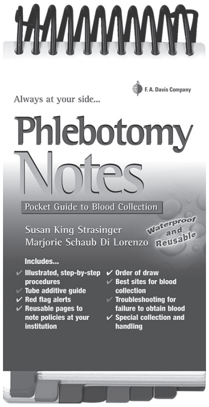
Molecular Diagnostics

Fundamentals, Methods, and Clinical Applications

THIRD EDITION

BOOST YOUR CONFIDENCE

ANYTIME | ANYWHERE


F. A. Davis Company
Always at your side...
Phlebotomy Notes
Pocket Guide to Blood Collection
Susan King Strasinger
Marjorie Schaub Di Lorenzo
Waterproof and Reusable
Includes...
✓ Illustrated, step-by-step procedures
✓ Tube additive guide
✓ Red flag alerts
✓ Reusable pages to note policies at your institution
✓ Order of draw
✓ Best sites for blood collection
✓ Troubleshooting for failure to obtain blood
✓ Special collection and handling

★★★★★

Awesome! *I have referred all my classmates to purchase this excellent resource. Even my instructor was impressed with it. You will not find a more handy and complete pocket guide.”*

★★★★★

Just what I needed. *It was exactly what I needed for class. It has helped through my clinical studies as well. Anytime I need an answer it has it for me.”*

—Amazon Reviewers

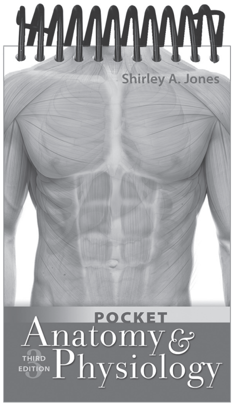
★★★★★

This book is great. *“This book helped me get an A in anatomy and physiology, this was my pocket book anywhere. I still use it after the class. I love it!”*

★★★★★

THE BEST!!! *“This is the best anatomy book I’ve ever seen!!! Don’t let the pocket size of it fool you! This anatomy guide is so detailed and complete!!! Very well done!!! Love it! Love it! Love it!!!”*

—Amazon Reviewers


Shirley A. Jones
POCKET
Anatomy & Physiology
THIRD EDITION

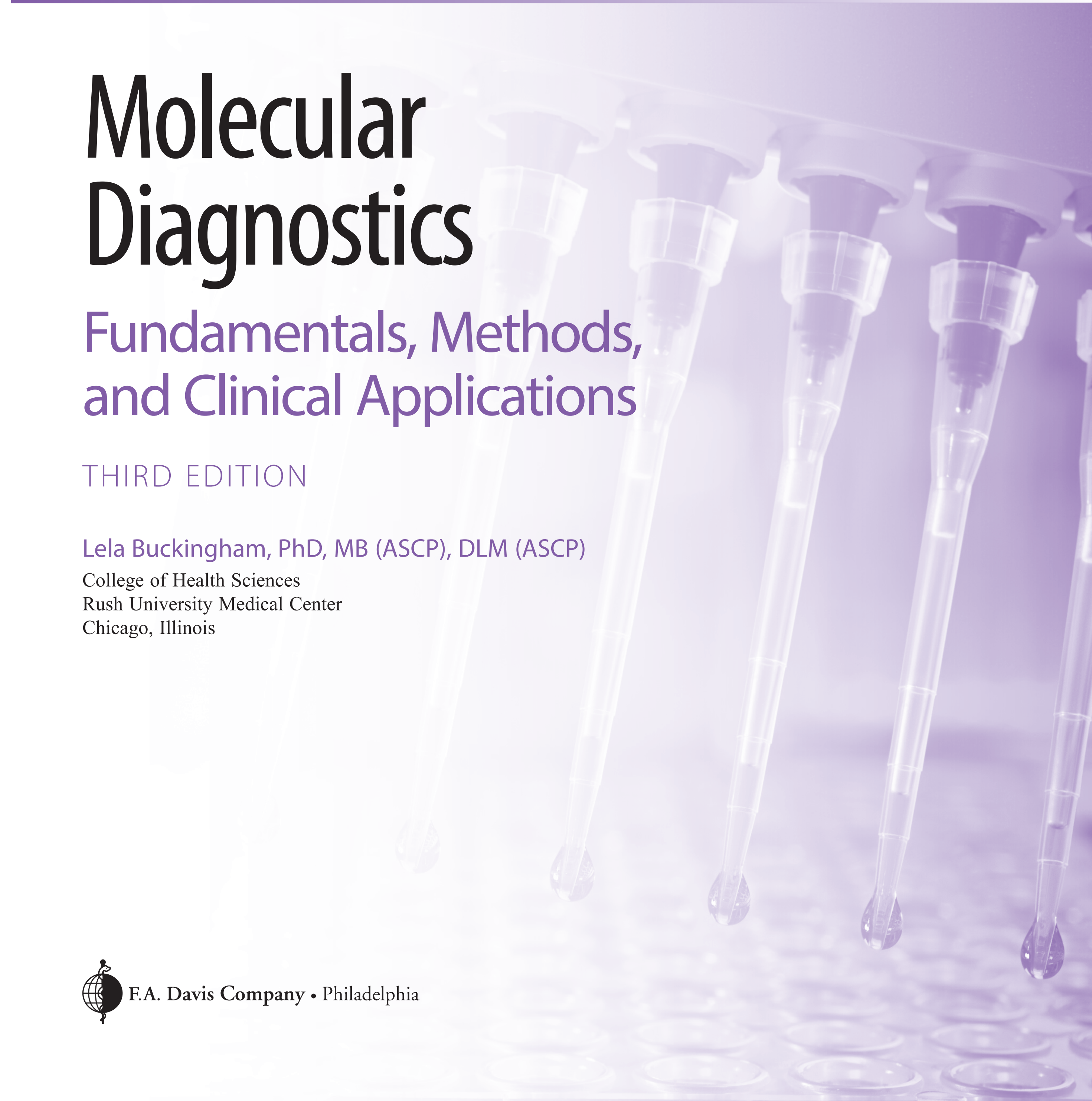
SAVE 20%
+ FREE SHIPPING

Use Promo Code: **DAVIS20**

ORDER TODAY!

FADavis.com

Promotion subject to change without notice. Offer valid for individual purchases from FADavis.com in the U.S. only.



Molecular Diagnostics

Fundamentals, Methods, and Clinical Applications

THIRD EDITION

Lela Buckingham, PhD, MB (ASCP), DLM (ASCP)
College of Health Sciences
Rush University Medical Center
Chicago, Illinois

F.A. Davis Company • Philadelphia

Acknowledgments

Thanks and gratitude are extended to all who helped in the completion of the third edition of this work. The useful input provided by reviewers who gave their valuable time to comment on and improve the writing is gratefully acknowledged. Molecular laboratory science is an example of innovative technology applied to the ultimate goal of improved patient care.

I owe thanks to my colleagues at Rush University Medical Center, Dr. Wei-Tong Hsu, Dr. Nick Moore, Dr. Mary Hayden, Dr. Sivadasan Kanangat, and Dr. Elizabeth Berry-Kravis, for help and support in their areas of expertise. I would also like to acknowledge colleagues in the Rush University College of Health Sciences, residents, fellows, students, and laboratory professionals who provided suggestions for the third edition, particularly Alexandra Vardouniotis, Dr. Mezgebe Gebrekiristos, and Adrian Tira, with whom I work and from whom I learn every day.

This book was originally envisioned by my co-author for the first edition, Dr. Maribeth Flaws. Thanks to her for initiating this project. Thanks to Dr. Herb Miller and the Medical Laboratory Science faculty in the Rush University College of Health Sciences for the opportunity to participate in medical laboratory science education. I greatly appreciate the guidance and support of the publication staff at F. A. Davis—Christa Fratantoro, Julie Chase, Roxanne Klaas, and Katharine Margeson—for the illustration and production of the text.

I would like to acknowledge and thank fellow members of the Association for Molecular Pathology, a vibrant and resourceful organization dedicated to education and policy in the practice of molecular diagnostics. This organization has provided an outlet for contextual information, training, and sanction to further this ever-advancing field of study.

F. A. Davis Company
1915 Arch Street
Philadelphia, PA 19103
www.fadavis.com

Copyright © 2019 by F. A. Davis Company

Copyright © 2019 by F. A. Davis Company. All rights reserved. This product is protected by copyright. No part of it may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without written permission from the publisher.

Printed in the United States of America

Last digit indicates print number: 10 9 8 7 6 5 4 3 2 1

Acquisitions Editor: Christa A. Fratantoro
Director of Content Development: George W. Lang
Content Project Manager: Julie Chase
Art and Design Manager: Carolyn O'Brien

As new scientific information becomes available through basic and clinical research, recommended treatments and drug therapies undergo changes. The author(s) and publisher have done everything possible to make this book accurate, up to date, and in accord with accepted standards at the time of publication. The author(s), editors, and publisher are not responsible for errors or omissions or for consequences from application of the book, and make no warranty, expressed or implied, in regard to the contents of the book. Any practice described in this book should be applied by the reader in accordance with professional standards of care used in regard to the unique circumstances that may apply in each situation. The reader is advised always to check product information (package inserts) for changes and new information regarding dose and contraindications before administering any drug. Caution is especially urged when using new or infrequently ordered drugs.

Library of Congress Cataloging-in-Publication Data

Names: Buckingham, Lela, author.
Title: Molecular diagnostics : fundamentals, methods, and clinical
applications / Lela Buckingham.
Description: Third edition. | Philadelphia : F.A. Davis Company, [2019] |
Includes bibliographical references and index.
Identifiers: LCCN 2018058583 (print) | LCCN 2018059084 (ebook) |
ISBN 9780803699540 | ISBN 9780803668294 (alk. paper)
Subjects: | MESH: Molecular Diagnostic Techniques—methods | Nucleic
Acids—analysis | Genetic Techniques
Classification: LCC RB43.7 (ebook) | LCC RB43.7 (print) | NLM QY 102 |
DDC 616.9/041—dc23
LC record available at https://lccn.loc.gov/2018058583

Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by F. A. Davis Company for users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that the fee of $.25 per copy is paid directly to CCC, 222 Rosewood Drive, Danvers, MA 01923. For those organizations that have been granted a photocopy license by CCC, a separate system of payment has been arranged. The fee code for users of the Transactional Reporting Service is: 978-0-8036-6829-4 / + $.25.

To our students committed to
service through the practice of Medical Laboratory Science

Preface

Molecular technology continues to grow in importance in the clinical laboratory. Training of health-care professionals routinely includes molecular biology, from laboratory techniques to therapeutic decisions. This textbook was written to provide fundamental knowledge of molecular biology, current methods, and their clinical applications.

The primary audience for this text is students enrolled in Clinical/Medical Laboratory Science programs at all levels. It explains the principles of molecular technology that are used for diagnostic purposes. Examples of applications of molecular-based assays are included in the text, along with case studies that illustrate the use and interpretation of these assays in patient care.

This text is also appropriate for those in other health-related disciplines who need to understand the purpose, principles, and interpretation of the molecular diagnostic tests that they will be ordering and assessing for their patients.

Students who are first learning about molecular-based assays will find the text useful for explaining the fundamental principles. Practitioners who are performing and interpreting these assays can use this text as a resource for reference and troubleshooting and to drive the implementation of additional molecular-based assays in their laboratories.

An Instructor's Resource package has been developed for educators who adopt this text for a course. These resources, including PowerPoint presentations, a test-item bank, and additional case studies, are available on DavisPlus at http://davisplus.fadavis.com.

Lela Buckingham

Reviewers

Catherine E. Bammert, MS, CT, MB (ASCP)^{CM}
Associate Professor
Program Director, Diagnostic Molecular Science
Clinical Laboratory Sciences
Northern Michigan University
Marquette, Michigan

Katie Bennett, PhD, MB (ASCP), NRCC-CC
Assistant Professor and Laboratory Director
Laboratory Sciences and Primary Care
Texas Tech University Health Sciences Center
Lubbock, Texas

Tammy Carter, PhD, MT (ASCP), MB (ASCP)
Assistant Professor, CLS Program Director
Laboratory Science and Primary Care
Texas Tech University Health Sciences Center
Lubbock, Texas

Kristen Coffey, MS
Visiting Instructor
Medical Laboratory Sciences
University of West Florida
Pensacola, Florida

Daniel Harrigan, MS, MB (ASCP)^{CM}
Professor
Laboratory Sciences
Blackhawk Technical College
Monroe, Wisconsin

Rachel Hulse, MS, MLS (ASCP)^{CM}
Program Director
Medical Laboratory Sciences
Idaho State University
Pocatello, Idaho

Marisa K. James, MA, MLS (ASCP)^{CM}
Program Director
School of Clinical Laboratory Science
North Kansas City Hospital
North Kansas City, Missouri

Jacqueline Peacock, PhD, MB (ASCP)^{CM}
Assistant Professor and Program Coordinator
Clinical Laboratory, Respiratory Care, and Health Administration Programs
Ferris State University
Grand Rapids, Michigan

David Petillo, PhD, MT (ASCP)^{CM}, MB
Clinical Coordinator/Assistant Professor
College of Health Professions
Molecular Diagnostics Program
Ferris State University
Grand Rapids, Michigan

Linda M. Ray, MS (ASCP)^{CM}
Assistant Professor
Medical Laboratory Science
University of North Dakota, School of Medicine & Health Sciences
Grand Forks, North Dakota

Barbara Sawyer, PhD, MLS (ASCP), MB
Professor
Lab Sciences and Primary Care
Texas Tech University Health Sciences Center
Lubbock, Texas

Ebot Sahidu Tabe, BMLS, MS, PhD, MB (ASCP)^{CM}
Instructor
Basic and Clinical Sciences
Albany College of Pharmacy and Health Sciences
Albany, New York

Geoffrey Toner, MS, MB (ASCP)^{CM}
Instructor/Education Coordinator
Medical Laboratory Sciences and Biotechnology
Jefferson College of Health Professions
Thomas Jefferson University
Philadelphia, Pennsylvania

EDITORIAL REVIEWERS

We especially thank our editorial reviewers for assisting with page proof review.

Mezgebe Gebrekiristos, PhD, MS, MT (ASCP)
Clinical Laboratory Scientist
Molecular Oncology Laboratory
Pathology Department
Rush University Medical Center
Chicago, Illinois

Lenny K. Hong, M.S., MLS (ASCP)^{CM}, MB (ASCP)^{CM}
PhD Graduate Student - Department of Pathology
University of Illinois at Chicago - College of Medicine
Chicago, Illinois

Adrian Tira, MSc., MT (ASCP)
Department of Pathology
Rush University Medical Center
Chicago, Illinois

Alexandra Vardouniotis MS, MLS (ASCP)^{CM} SBB^{CM}
Medical Laboratory Scientist
Rush University Medical Center
Chicago, Illinois

Contents

Section I

Fundamentals of Molecular Biology: An Overview 1

1 Nucleic Acids and Proteins 2
- DNA 3
- DNA STRUCTURE 4
  - *Nucleotides 4*
  - *Nucleic Acid 8*
- DNA REPLICATION 9
  - *Polymerases 11*
- ENZYMES THAT METABOLIZE DNA 14
  - *Restriction Enzymes 14*
  - *DNA Ligase 17*
  - *Other DNA Metabolizing Enzymes 17*
- RECOMBINATION IN SEXUALLY REPRODUCING ORGANISMS 19
- RECOMBINATION IN ASEXUAL REPRODUCTION 21
  - *Conjugation 21*
  - *Transduction 23*
  - *Transformation 23*
- PLASMIDS 25
- RNA 27
  - *Transcription 27*
  - *Transcription Initiation 28*
  - *Transcription Elongation 28*
  - *Transcription Termination 29*
- TYPES/STRUCTURES OF RNA 29
  - *Ribosomal RNA 29*
  - *Messenger RNA 29*
  - *Small Nuclear RNA 33*
  - *Transfer RNA 33*
  - *Other RNAs 35*
- RNA POLYMERASES 35
- OTHER RNA-METABOLIZING ENZYMES 36
  - *Ribonucleases 36*
  - *RNA Helicases 37*
- PROTEINS AND THE GENETIC CODE 37
  - *Amino Acids 38*
  - *Genes 43*
  - *The Genetic Code 43*
- TRANSLATION 46
  - *Amino Acid Charging 46*
  - *Protein Synthesis 47*

2 Gene Expression and Epigenetics 57
- TRANSCRIPTION 58
  - *Transcription Initiation 58*
  - *Transcription Elongation 58*
  - *Transcription Termination 59*
- REGULATION OF TRANSCRIPTION 60
  - *Regulation of Messenger RNA Synthesis at Initiation 60*
  - *Post-Transcriptional Regulation 64*
  - *Post-Translational Regulation 64*
- EPIGENETICS 65
  - *Histone Modification 65*
  - *Nucleic Acid Methylation 66*
- CLASSIFICATION OF EPIGENETIC FACTORS 69
- NONCODING RNAs 69
  - *MicroRNAs 70*
  - *Small Interfering RNAs 70*
  - *Other Small RNAs 70*
  - *Long Noncoding RNAs 71*

Section II

Common Techniques in Molecular Biology 77

3 Nucleic Acid Extraction Methods 78
  ISOLATION OF DNA 79
    Preparing the Sample 79
    DNA Isolation Chemistries 81
  ISOLATION OF RNA 87
    Total RNA 87
    Specimen Collection 87
    RNA Isolation Chemistries 87
  MEASUREMENT OF NUCLEIC ACID QUALITY AND QUANTITY 90
    Electrophoresis 90
    Spectrophotometry 92
    Fluorometry 93
    Microfluidics 94

4 Resolution and Detection of Nucleic Acids 97
  ELECTROPHORESIS OF NUCLEIC ACIDS 98
  GEL SYSTEMS 99
    Agarose Gels 99
    Polyacrylamide Gels 101
  CAPILLARY ELECTROPHORESIS 102
  BUFFER SYSTEMS 104
    Buffer Additives 105
  ELECTROPHORESIS EQUIPMENT 106
    Gel Loading 108
  DETECTION SYSTEMS 109
    Fluorescent Dyes 109
    Silver Stain 110

5 Analysis and Characterization of Nucleic Acids and Proteins 112
  RESTRICTION ENZYME MAPPING OF DNA 113
  CRISPR ENZYME SYSTEMS 115
  HYBRIDIZATION TECHNOLOGIES 116
    Southern Blots 117
  PROBE HYBRIDIZATION 121
    Northern Blots 122
    Western Blots 122
  PROBES 123
    DNA Probes 123
    RNA Probes 124
    Other Nucleic Acid Probe Types 124
    Protein Probes 124
    Probe Labeling 126
    Nucleic Acid Probe Design 126
  HYBRIDIZATION CONDITIONS, STRINGENCY 128
  DETECTION SYSTEMS 129
  INTERPRETATION OF RESULTS 132
  ARRAY-BASED HYBRIDIZATION 133
    Dot/Slot Blots 133
    Genomic Array Technology 134
  SOLUTION HYBRIDIZATION 138

6 Nucleic Acid Amplification 142
  TARGET AMPLIFICATION 143
    Polymerase Chain Reaction 143
    Transcription-Based Amplification Systems 164
    Genomic Amplification Methods 165
  PROBE AMPLIFICATION 168
    Ligase Chain Reaction 168
    Strand Displacement Amplification 169
    Qβ Replicase 170
  SIGNAL AMPLIFICATION 172
    Branched DNA Amplification 172
    Hybrid Capture Assays 173
    Cleavage-Based Amplification 173
    Cycling Probe 174

7 Chromosomal Structure and Chromosomal Mutations 179
  CHROMOSOMAL STRUCTURE AND ANALYSIS 181
    Chromosomal Compaction and Histones 181
    Chromosome Morphology 183
    Visualizing Chromosomes 184
  DETECTION OF GENOME AND CHROMOSOMAL MUTATIONS 186
    Karyotyping 186
    Fluorescence In Situ Hybridization 190
  COMPARATIVE GENOME HYBRIDIZATION (CGH) 195

8 Gene Mutations 199
  TYPES OF GENE MUTATIONS 200
  DETECTION OF GENE MUTATIONS 201
    Biochemical Methods 201
    Nucleic Acid Analyses 207
  GENE VARIANT NOMENCLATURE 218
  GENE NAMES 219

9 DNA Sequencing 223
  DIRECT SEQUENCING 224
    Manual Sequencing 224
    Automated Fluorescent Sequencing 229

PYROSEQUENCING 235
BISULFITE DNA SEQUENCING 236
RNA SEQUENCING 237
NEXT-GENERATION SEQUENCING 238
Gene Panels 240
NGS Library Preparation 240
Targeted Libraries 241
Sequencing Platforms 243
Sequence Quality 246
Filtering and Annotation 247
BIOINFORMATICS 248
THE HUMAN GENOME PROJECT 250
Variant Associations With Phenotype 253

Section III

Techniques in the Clinical Laboratory 259

10 DNA Polymorphisms and Human Identification 260
TYPES OF POLYMORPHISMS 261
RFLP TYPING 262
Genetic Mapping With RFLPs 264
RFLP and Parentage Testing 264
Human Identification Using RFLPs 265
STR TYPING BY PCR 266
STR Analysis 268
Y-STR 278
LINKAGE ANALYSIS 282
BONE MARROW ENGRAFTMENT TESTING USING DNA POLYMORPHISMS 283
PSTR Testing 285
Post-Transplant Engraftment Testing 287
QUALITY ASSURANCE FOR SURGICAL SECTIONS USING STR 289
SINGLE-NUCLEOTIDE POLYMORPHISMS 289
The Human Haplotype Mapping (HapMap) Project 290
MITOCHONDRIAL DNA POLYMORPHISMS 291
OTHER IDENTIFICATION METHODS 293
Protein-Based Identification 293
Epigenetic Profiles 294

11 Detection and Identification of Microorganisms 301
SPECIMEN COLLECTION 302
SAMPLE PREPARATION 304
QUALITY ASSURANCE 305
Controls 305
Quality Control 305
Selection of Sequence Targets for Detection of Microorganisms 307
MOLECULAR DETECTION OF MICROORGANISMS 308
Bacteria 309
Viruses 313
Mycology 323
Parasites 324
ANTIMICROBIAL AGENTS 324
Resistance to Antimicrobial Agents 325
Molecular Detection of Resistance 327
MOLECULAR EPIDEMIOLOGY 329
Molecular Strain Typing Methods for Epidemiological Studies 330
Comparison of Typing Methods 336

12 Molecular Detection of Inherited Diseases 344
THE MOLECULAR BASIS OF INHERITED DISEASES 345
CHROMOSOMAL ABNORMALITIES 345
PATTERNS OF INHERITANCE IN SINGLE-GENE DISORDERS 346
MOLECULAR BASIS OF SINGLE-GENE DISORDERS 349
Lysosomal Storage Diseases 349
Factor V Leiden 349
Prothrombin 349
Methylenetetrahydrofolate Reductase 352
Hemochromatosis 352
Cystic Fibrosis 353
Cytochrome P-450 354
SINGLE-GENE DISORDERS WITH NONCLASSICAL PATTERNS OF INHERITANCE 355
Mutations in Mitochondrial Genes 356
Nucleotide-Repeat Expansion Disorders 357
Genomic Imprinting 362
Multifactorial Inheritance 363
LIMITATIONS OF MOLECULAR TESTING 363

13 Molecular Oncology 369
CLASSIFICATION OF NEOPLASMS 370
MOLECULAR BASIS OF CANCER 371
ANALYTICAL TARGETS OF MOLECULAR TESTING 372
GENE AND CHROMOSOMAL MUTATIONS IN SOLID TUMORS 372
Human Epidermal Growth Factor Receptor 2, HER2/neu/erb-b2 1 *(17q21.1) 372*
Epidermal Growth Factor Receptor, EGFR *(7p12) 373*

  - *Kirsten Rat Sarcoma Viral Oncogene Homolog,* K-ras *(12p12); Neuroblastoma ras,* N-ras *(1p13); and Harvey Rat Sarcoma Viral Oncogene Homolog,* H-ras *(11p15) 375*
  - *Ewing Sarcoma,* EWS *(22q12) 376*
  - *Synovial Sarcoma Translocation, Chromosome 18—Synovial Sarcoma Breakpoint 1 and 2,* SYT-SSX1, SYT-SSX2 *t(X;18)(p11.2;q11.2) 377*
  - *Paired Box-Forkhead in Rhabdomyosarcoma,* PAX3-FKHR, PAX7-FKHR, *t(1;13), t(2;13) 378*
  - *Tumor Protein 53,* TP53 *(17p13) 378*
  - *Ataxia Telangiectasia Mutated Gene,* ATM *(11q22) 379*
  - *Breast Cancer 1 Gene,* BRCA1 *(17q21), and Breast Cancer 2 Gene,* BRCA2 *(13q12) 380*
  - *Von Hippel–Lindau Gene,* VHL *(3p26) 380*
  - *V-myc Avian Myelocytomatosis Viral-Related Oncogene, Neuroblastoma-Derived,* MYCN *or* n-myc *(2p24) 381*
  - *V-Ros Avian UR2 Sarcoma Virus Oncogene Homolog 1* (ROS1) *Proto-Oncogene (6q22.1) and Rearranged During Transfection* (RET) *Proto-Oncogene (10q11) 381*
  - *Anaplastic Lymphoma Receptor Tyrosine Kinase* (ALK) *Proto-Oncogene, 2p23.1 382*
  - *V-Kit Hardy-Zuckerman 4 Feline Sarcoma Viral Oncogene Homolog,* KIT, c-KIT *(4q12) 382*
  - *Other Molecular Abnormalities 382*
  - *Microsatellite Instability 382*
  - *Loss of Heterozygosity 385*
  - *Liquid Biopsy 386*
- MOLECULAR ANALYSIS OF LEUKEMIA AND LYMPHOMA 387
  - *Gene Rearrangements 387*
  - *Mutations in Hematological Malignancies 397*
  - *Mutation Spectra 405*

14 DNA-Based Tissue Typing 417

- THE MHC LOCUS 418
- HLA POLYMORPHISMS 420
  - *HLA Nomenclature 420*
- MOLECULAR ANALYSIS OF THE MHC 425
  - *Serological Analysis 427*
  - *DNA-Based Typing 430*
  - *Combining Typing Results 436*
  - *HLA Test Discrepancies 436*
  - *Coordination of HLA Test Methods 437*
- ADDITIONAL RECOGNITION FACTORS 437
  - *Minor Histocompatibility Antigens 437*
  - *Nonconventional MHC Antigens 437*
  - *Killer Cell Immunoglobulin-Like Receptors 437*
- MHC DISEASE ASSOCIATION 438
- SUMMARY OF LABORATORY TESTING 439

15 Quality Assurance and Quality Control in the Molecular Laboratory 446

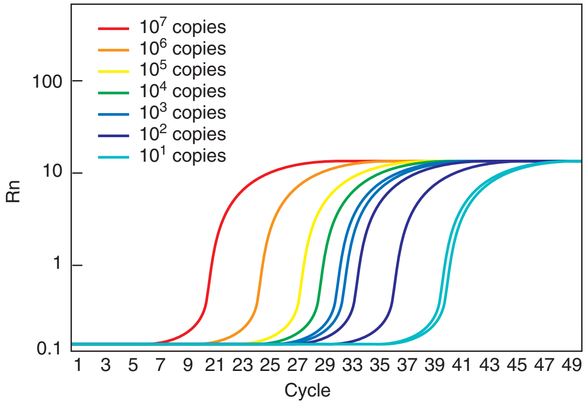
- SPECIMEN HANDLING 447
  - *Collection Tubes for Molecular Testing 448*
  - *Precautions 450*
  - *Holding and Storage Requirements 451*
- TEST PERFORMANCE 451
  - *Next-Generation Sequencing 456*
  - *Calibrators and Method Calibration 456*
  - *Controls 457*
- QUALITY CONTROL 458
- QUALITY ASSURANCE 458
- INSTRUMENT MAINTENANCE 459
  - *Instrument Calibration 463*
- REAGENTS 463
  - *Reagent Categories 464*
  - *Chemical Safety 465*
  - *Reagent Storage 466*
  - *Reagent Labeling 466*
- PROFICIENCY TESTING 468
- DOCUMENTATION OF TEST RESULTS 468
  - *Gene Nomenclature 469*
  - *Gene Sequencing Results 469*
- REPORTING RESULTS 469

Appendix A Study Question Answers 473

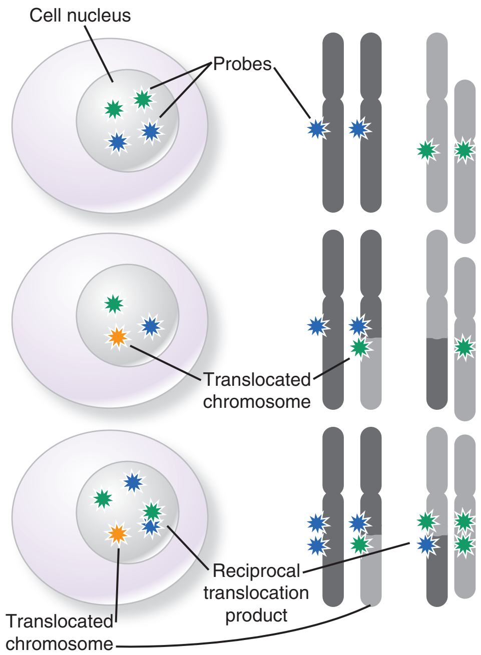
Appendix B Answers to Case Studies 501

Glossary 505

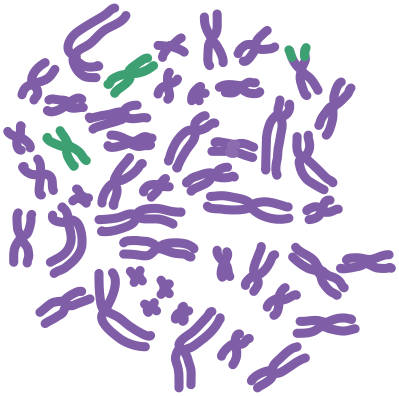
Index 529



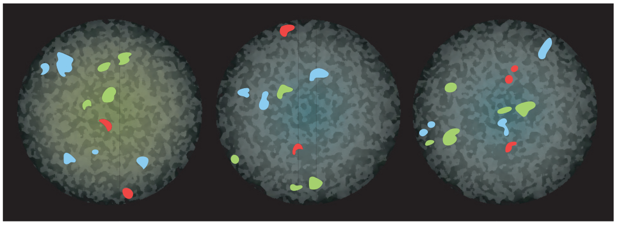
COLOR PLATE 1 A plot of the accumulation of polymerase chain reaction (PCR) product over 50 cycles of PCR. In this sigmoid curve, the generation of fluorescence occurs earlier with more starting template (solid lines) than with less (dotted lines). See Figure 6.13A in the text.



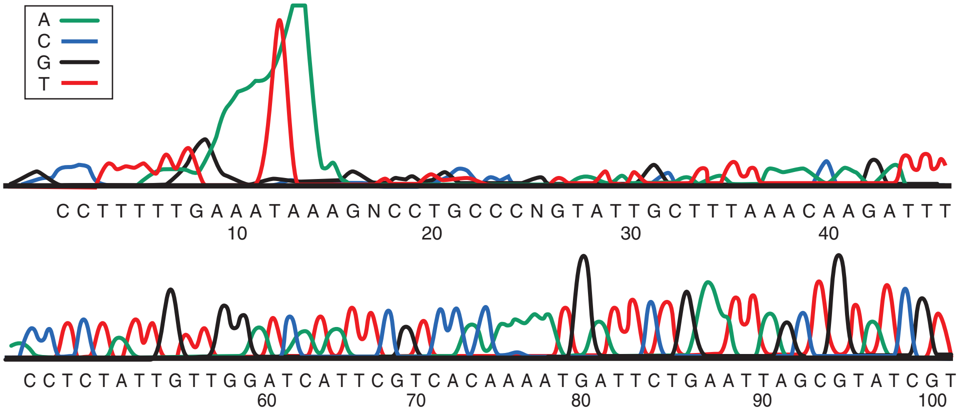
COLOR PLATE 2 Fluorescence in situ hybridization (FISH) analysis using distinct probes to detect a translocation. A normal nucleus has two signals from each probe *(top)*. A translocation involving the two chromosomes combines the two probe colors *(middle)*. Dual-fusion probes confirm the presence of the translocation by also giving a signal from the reciprocal breakpoint *(bottom)*. See Figure 7.16 in the text.



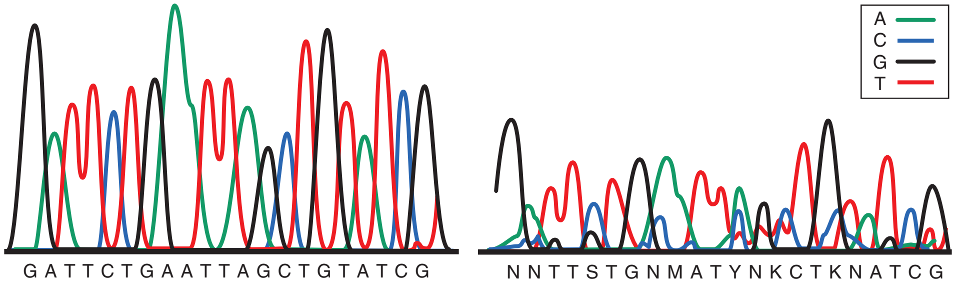
COLOR PLATE 3 Chromosome painting showing a derivative chromosome formed by the movement of a fragment of chromosome 12 (black) to an unidentified chromosome. See Figure 7.19 in the text.



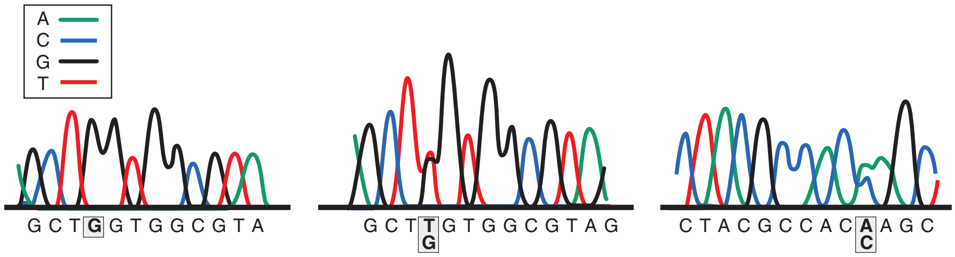
COLOR PLATE 4 Multicolor fluorescence in situ hybridization (FISH) analysis simultaneously reveals structural or numerical abnormalities in three loci. See Figure 7.21 in the text.



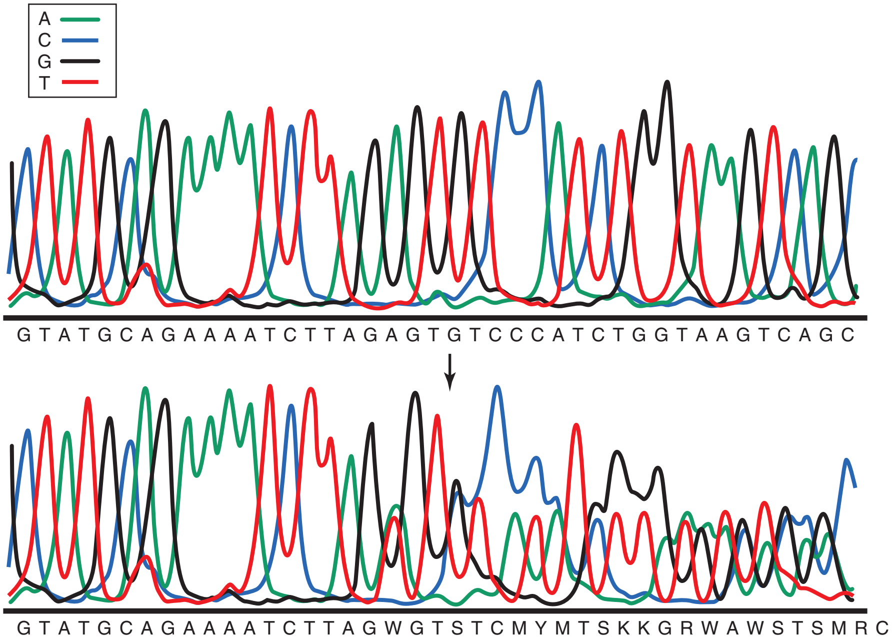
COLOR PLATE 5 Electropherogram showing a dye blob at the beginning of a sequence (nucleotide positions 9 to 15). The sequence read around this area is not accurate. See Figure 9.10 in the text.



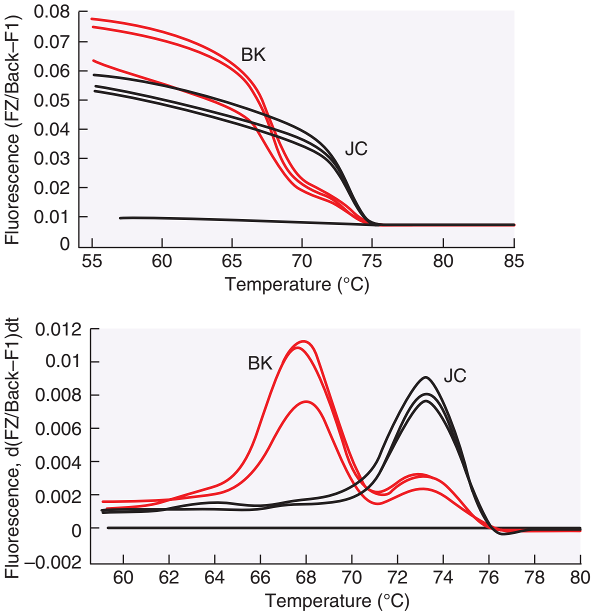
COLOR PLATE 6 Examples of good sequence quality *(left)* and poor sequence quality *(right)*. Note the clean baseline on the good sequence; that is, only one color peak is present at each nucleotide position. Automatic sequence-reading software will not accurately call a poor sequence. Compare the text sequences below the two scans. See Figure 9.11 in the text.



COLOR PLATE 7 Sequencing of a heterozygous G to T mutation in exon 12 of the *KRAS* gene. The normal codon sequence is GGT *(left)*. The heterozygous mutation (GT; *center*) is confirmed in the reverse sequence (CA; *right*). See Figure 9.12 in the text.



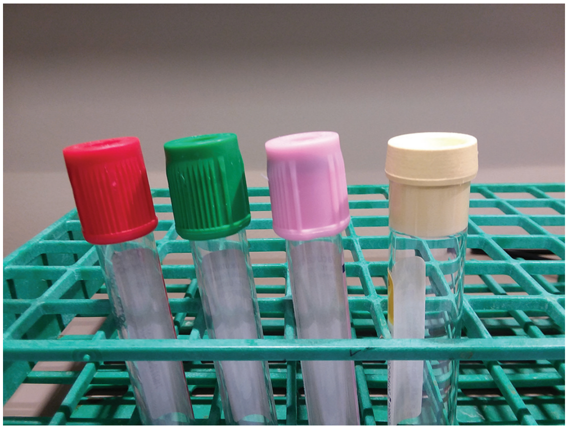
COLOR PLATE 8 The 187 delAG mutation in the *BRCA*1 gene detected by Sanger sequencing. This heterozygous dinucleotide deletion is evident in the lower panel where, at the site of the mutation, two sequences are overlaid: the normal sequence and the normal sequence minus two bases. See Figure 9.13 in the text.



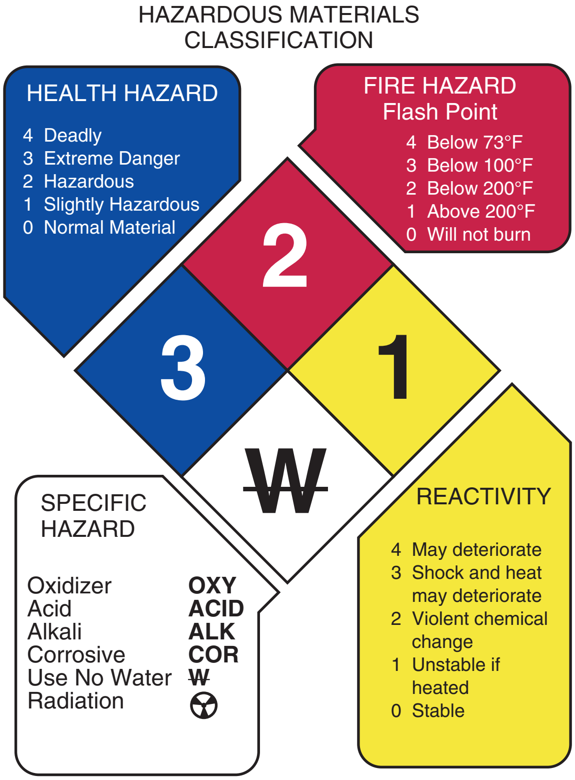
COLOR PLATE 9 Melt-curve analysis of BK and JC viruses. BK and JC are differentiated from one another by differences in the T_m* of the probe specific for each viral sequence. Fluorescence from double-stranded DNA decreases with increasing temperature and DNA denaturation to single strands *(top panel)*. Instrument software will present the derivative of the fluorescence *(bottom panel)* where the melting temperatures (T_m; 67°C to 68°C for BK and 73°C to 74°C for JC) are observed as peaks. See Figure 11.4 in the text.

WARNING
ATTENTION
CAUTION
RADIOACTIVE MATERIALS

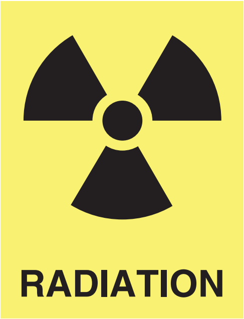
COLOR PLATE 10 Biohazard stickers are required for cabinets, refrigerators, or freezers that contain potentially hazardous reagents or patient specimens. See Figure 15.1 in the text.



COLOR PLATE 11 For molecular analysis, blood or bone marrow specimens collected in ethylenediaminetetraacetic acid (EDTA; lavender-cap) or acid citrate dextrose (ACD; yellow-cap) tubes are preferred. Heparin (green cap) is used for cytogenetic tests. Immunoassays or mass-spectrometry methods may be performed on serum collected in tubes without coagulant (red-cap tubes). See Figure 15.3 in the text.


HAZARDOUS MATERIALS CLASSIFICATION
HEALTH HAZARD
4 Deadly
3 Extreme Danger
2 Hazardous
1 Slightly Hazardous
0 Normal Material
FIRE HAZARD
Flash Point
4 Below 73°F
3 Below 100°F
2 Below 200°F
1 Above 200°F
0 Will not burn
2
3
1
W
SPECIFIC HAZARD
Oxidizer OXY
Acid ACID
Alkali ALK
Corrosive COR
Use No Water W
Radiation
REACTIVITY
4 May deteriorate
3 Shock and heat may deteriorate
2 Violent chemical change
1 Unstable if heated
0 Stable

COLOR PLATE 12 National Fire Protection Association (NFPA) hazard labels have three parts, labeled with numbers 0 to 4, depending on the severity of the hazard, from none (0) to severe (4). The fourth section has two categories. OXY indicates a strong oxidizer, which greatly increases the rate of combustion. The W symbol indicates dangerous reactivity with water, which would prohibit the use of water to extinguish a fire in the presence of this chemical. See Figure 15.17 in the text.



COLOR PLATE 13 Rooms, cabinets, and equipment containing radioactive chemicals are identified with radiation safety labels. See Figure 15.18 in the text.

Section I

Fundamentals of Molecular Biology: An Overview

Chapter 1

Nucleic Acids and Proteins

Outline

- **DNA**
- **DNA STRUCTURE**
  - *Nucleotides*
  - *Nucleic Acid*
- **DNA REPLICATION**
  - *Polymerases*
- **ENZYMES THAT METABOLIZE DNA**
  - *Restriction Enzymes*
  - *DNA Ligase*
  - *Other DNA Metabolizing Enzymes*
- **RECOMBINATION IN SEXUALLY REPRODUCING ORGANISMS**
- **RECOMBINATION IN ASEXUAL REPRODUCTION**
  - *Conjugation*
  - *Transduction*
  - *Transformation*
- **PLASMIDS**
- **RNA**
  - *Transcription*
  - *Transcription Initiation*
  - *Transcription Elongation*
  - *Transcription Termination*
- **TYPES/STRUCTURES OF RNA**
  - *Ribosomal RNA*
  - *Messenger RNA*
    - Messenger RNA Processing
  - *Small Nuclear RNA*
  - *Transfer RNA*
  - *Other RNAs*
- **RNA POLYMERASES**
- **OTHER RNA-METABOLIZING ENZYMES**
  - *Ribonucleases*
  - *RNA Helicases*
- **PROTEINS AND THE GENETIC CODE**
  - *Amino Acids*
  - *Genes*
  - *The Genetic Code*
- **TRANSLATION**
  - *Amino Acid Charging*
  - *Protein Synthesis*

Objectives

1.1 Diagram the structure of nitrogen bases, nucleosides, and nucleotides.

1.2 Describe the nucleic acid structure as a polymer of nucleotides.

1.3 Demonstrate how deoxyribonucleic acid (DNA) is replicated such that the order or sequence of nucleotides is maintained (semiconservative replication).

1.4 Relate how ribonucleic acid (RNA) is synthesized (transcription) compared with DNA

1.5 List and describe types of RNA.

1.6 Explain the reaction catalyzed by polymerases that results in the phosphodiester backbone of the nucleic acid chains.
1.7 Note how the replicative process results in the antiparallel nature of complementary strands of DNA.
1.8 List the enzymes that modify DNA and RNA, and state their specific functions.
1.9 Explain mRNA processing, including capping, polyadenylation, and splicing.
1.10 Illustrate three ways in which DNA can be transferred between bacterial cells.
1.11 Define recombination, and sketch how new combinations of genes are generated in sexual and asexual reproduction.
1.12 Outline the structure and chemical nature of the 20 amino acids.
1.13 Show how the chemistry of the amino acids affects their chemical characteristics.
1.14 Give the definition of a gene.
1.15 Recount how the genetic code was solved.
1.16 Describe how amino acids are polymerized into proteins, using RNA as a guide (translation).
1.17 Relate protein function to the structural domains of the amino acid sequence.

When James Watson coined the term *molecular biology,*[1] he was referring to the biology of deoxyribonucleic acid (DNA). Of course, there are other molecules in nature. The term, however, is still used to describe the study of nucleic acids. In the medical laboratory, molecular techniques are designed for the handling and analysis of the nucleic acids, DNA and ribonucleic acid (RNA). Protein analysis and that of carbohydrates and other molecular species might also be categorized as "molecular" studies performed by flow cytometry, in situ histology, and tissue typing. The molecular biology laboratory, therefore, may be a separate entity or part of an existing clinical pathology unit. This chapter will address the nucleic acids.

Nucleic acids offer several characteristics that support their use for clinical purposes. Highly specific analyses can be carried out without the requirement for extensive physical or chemical selection of target molecules or organisms, allowing specific and rapid analysis from limited specimens. Furthermore, information carried in the order or sequence of the nucleotides that make up the nucleic acids is the basis for normal and pathological **traits** from microorganisms to humans and, as such, provides a powerful means of predictive analysis. Effective prevention and treatment of disease will result from the analysis of these molecules in the medical laboratory.

DNA

Deoxyribonucleic acid (DNA) is a macromolecule of carbon, nitrogen, oxygen, phosphorous, and hydrogen atoms. It is assembled in units of **nucleotides** that are composed of a phosphorylated ribose sugar and a **nitrogen base.** There are four nitrogen bases that make up the majority of DNA found in all organisms. These are **adenine, cytosine, guanine,** and **thymine.** Nitrogen bases are attached to a **deoxyribose** sugar, which forms a polymer with the deoxyribose sugars of other nucleotides through a **phosphodiester bond.** Linear assembly of the nucleotides makes up one strand of DNA. Two strands of DNA comprise the DNA double helix.

In 1871, Johann Friedrich Miescher published a paper on **nuclein,** the viscous substance extracted from cell nuclei. In his writings, he made no mention of the function of nuclein. Walther Flemming, a leading cell biologist, describing his work on the nucleus in 1882 admitted that the biological significance of the substance was unknown. We now know that the purpose of DNA, contained in the nucleus of the cell, is to store information. The information in the DNA storage system is based on the order or sequence of nucleotides in the nucleic acid polymer. Just as computer information storage is based on sequences of 0 and 1, biological information is based on sequences of A, C, G, and T. These four building blocks (with a few modifications) account for all of the biological diversity that makes up life on earth.

Historical Highlights

Johann Friedrich Miescher is credited with the discovery of DNA in 1869.[2] Miescher had isolated white blood cells out of seepage collected from discarded surgical bandages. He found that he could extract a viscous substance from the cells in this material. Miescher also observed that most of the nonnuclear cell components could be lysed away with dilute hydrochloric acid, leaving the nuclei intact. Addition of extract of pig stomach (a source of pepsin to dissolve away contaminating proteins) resulted in a somewhat shrunken but clean preparation of nuclei. Extraction of these with alkali yielded the same substance isolated from the intact cells. It precipitated upon the addition of acid and redissolved in alkali. Chemical analysis of this substance demonstrated that it was 14% nitrogen and 2.5% phosphorus, different from any then-known group of biochemicals. He named the substance "nuclein." (Analytical data indicate that less than 30% of Miescher's first nuclein preparation was actually DNA.) He later isolated a similar viscous material from salmon sperm and noted: "If one wants to assume that a single substance . . . is the specific cause of fertilization, then one should undoubtedly first of all think of nuclein."

DNA STRUCTURE

The double helical structure of DNA (Fig. 1.1) was first described by James Watson and Francis Crick. Their molecular model was founded on previous observations of the chemical nature of DNA and physical evidence including diffraction analyses performed by Rosalind Franklin.[3] The helical structure of DNA results from the physicochemical demands of the linear array of nucleotides. Both the specific **sequence** (order) of nucleotides in the strand, as well as the surrounding chemical microenvironment, can affect the nature of the DNA helix.

Nucleotides

The four nucleotide building blocks of DNA are molecules of about 700 kd. Each nucleotide consists of a five-carbon sugar, the first carbon of which is covalently joined to a nitrogen base and the fifth carbon to a phosphate moiety (Fig. 1.2). A nitrogen base bound to an unphosphorylated sugar is a **nucleoside. Adenosine (A), guanosine (G), cytidine (C),** and **thymidine (T)** are nucleosides. If the ribose sugar is phosphorylated, the molecule is a nucleoside mono-, di-, or triphosphate or a nucleotide. For example, adenosine with one phosphate is adenosine monophosphate (AMP). Adenosine with three phosphates is adenosine triphosphate (ATP). Free nucleotides are deoxyribonucleoside triphosphates (e.g., dATP). They are routinely designated as A, C, G, and T in the DNA molecule. Nucleotides can be converted to nucleosides by hydrolysis.

The five-carbon sugar of DNA is deoxyribose, which is ribose with the number-two carbon of deoxyribose linked to a hydrogen atom rather than a hydroxyl group (see Fig. 1.2). The hydroxyl group on the third carbon is important for forming the phosphodiester bond that is the backbone of the DNA strand.

Nitrogen bases are planar carbon-nitrogen ring structures. The four common nitrogen bases in DNA are adenine, guanine, cytosine, and thymine. Amine and ketone substitutions, as well as the single or double bonds within the rings, distinguish the four bases that comprise the majority of DNA (Fig. 1.3). Nitrogen bases with a single-ring structure (thymine, cytosine) are **pyrimidines.** Bases with a double-ring structure (guanine, adenine) are **purines.**

The numbering of the positions in the nucleotide molecule starts with the ring positions of the nitrogen base, designated C or N 1, 2, 3, and so on. The carbons of the ribose sugar are numbered 1′ to 5′, distinguishing the positions of the sugar rings from those of the nitrogen base rings (Fig. 1.4).

Advanced Concepts

The double helix first described by Watson and Crick is DNA in its hydrated form (B-form) and is the standard form of DNA.[4] It has 10.5 steps or pairs of nucleotides (base pairs [bp]) per turn. Dehydrated DNA takes the A-form with about 11 bp per turn and the center of symmetry along the outside of the helix rather than down the middle as it is in the B-form. Both A- and B-form DNA are right-handed helices. Stress and torsion can throw the double helix into a Z-form. Z-DNA

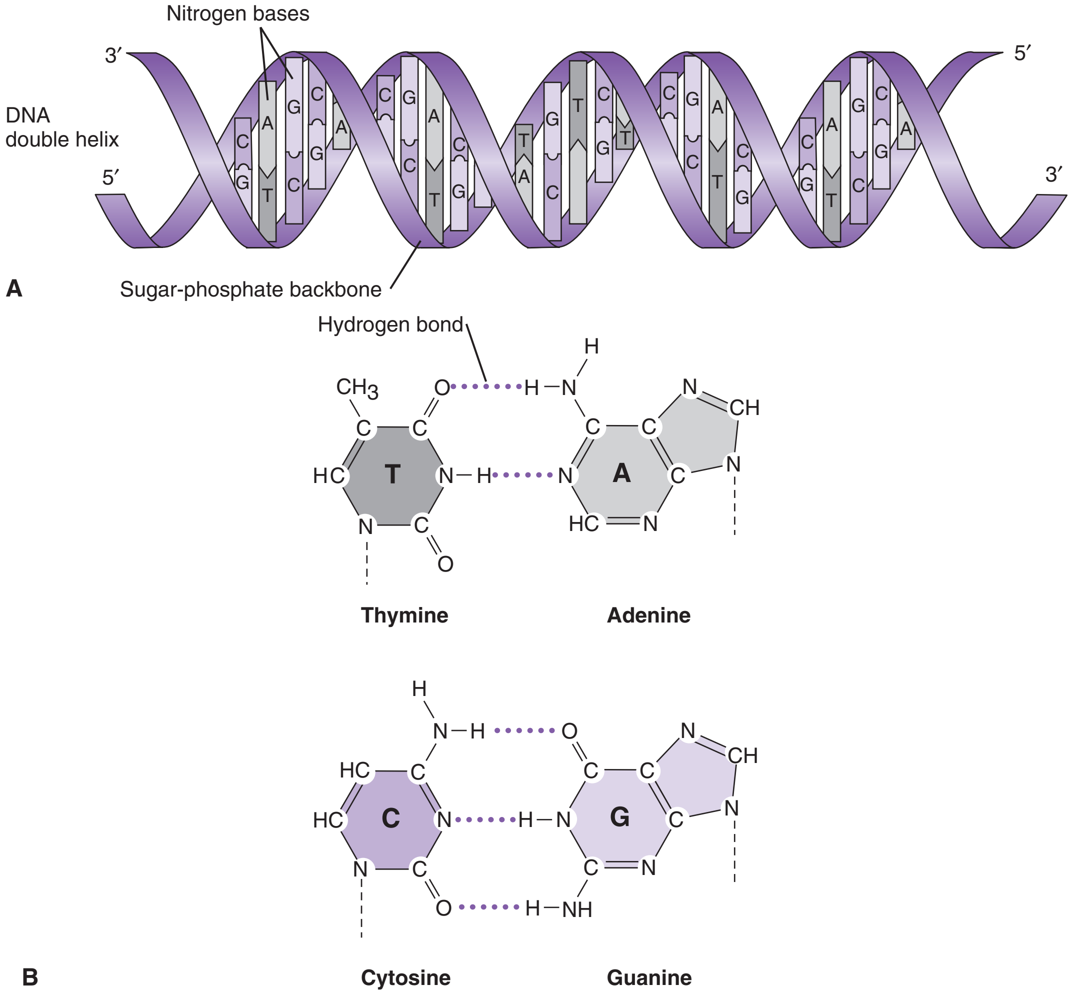

Nitrogen bases
3′
5′
DNA double helix
5′
3′
A
Sugar-phosphate backbone
Hydrogen bond
Thymine
Adenine
B
Cytosine
Guanine

FIGURE 1.1 (A) The double helix. The phosphodiester backbones of the two nucleic acid chains form the helix. Nitrogen bases are oriented toward the center, where they hydrogen bond with homologous bases to stabilize the structure. (B) Two hydrogen bonds form between adenine and thymine. Three hydrogen bonds form between guanine and cytosine.

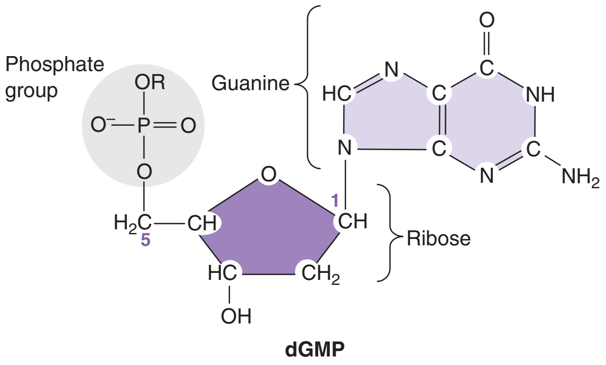

Phosphate group
Guanine
Ribose
1
5
dGMP

FIGURE 1.2 The nucleotide deoxyguanosine 5′ phosphate or guanosine monophosphate (dGMP). It is composed of deoxyribose covalently bound at its number 1 carbon to the nitrogen base, guanine, and at its number 5 carbon to a phosphate group. The molecule without the phosphate group is the nucleoside, deoxyguanosine.

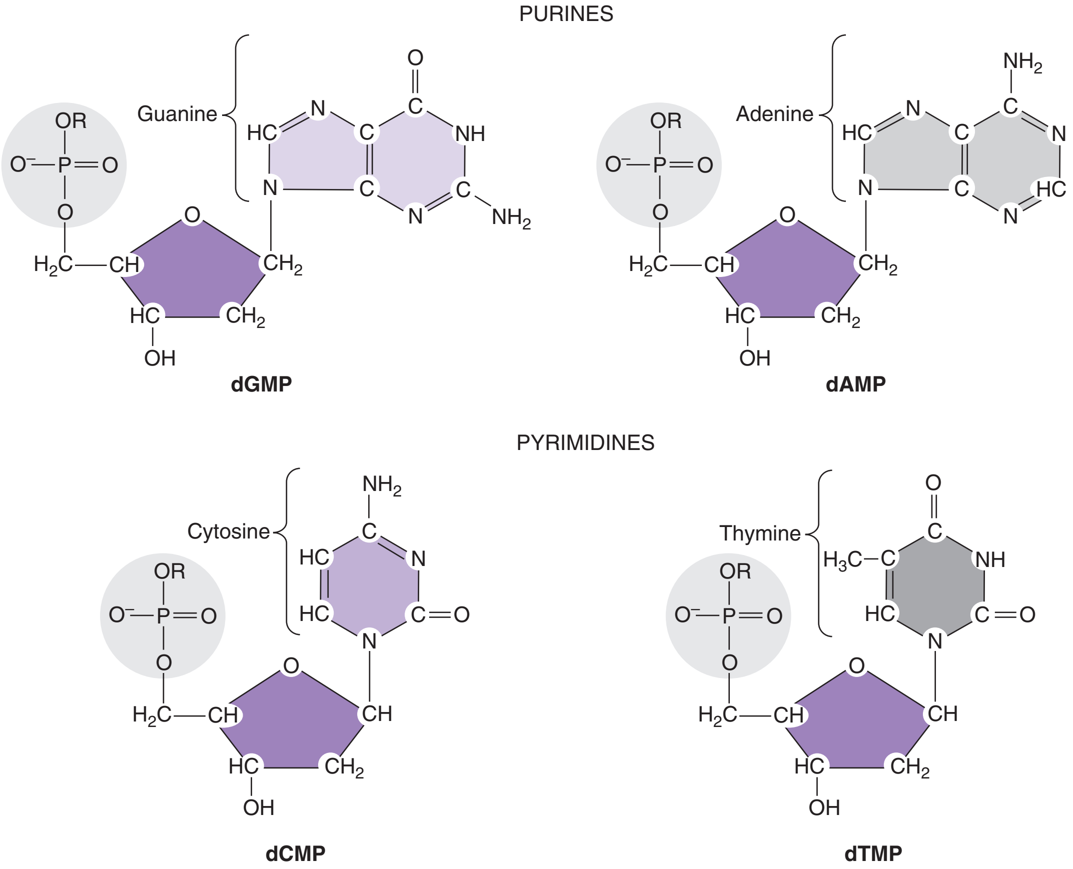

PURINES
Guanine
Adenine
OR
O⁻—P=O
O
H_2C—CH
HC
OH
CH_2
HC=N
N
C
O
NH
NH_2
dGMP
dAMP
PYRIMIDINES
Cytosine
Thymine
H_3C—C
C=O
CH
dCMP
dTMP

FIGURE 1.3 Nucleotides, deoxyguanosine monophosphate (dGMP), deoxyadenosine monophosphate (dAMP), deoxythymidine monophosphate (dTMP), and deoxycytidine monophosphate (dCMP), differ by the attached nitrogen bases. The nitrogen bases, guanine and adenine, have purine ring structures. Thymine and cytosine have pyrimidine ring structures. Uracil, the nucleotide base that replaces thymine in RNA, has the pyrimidine ring structure of thymine minus the methyl group and hydrogen bonds with adenine.

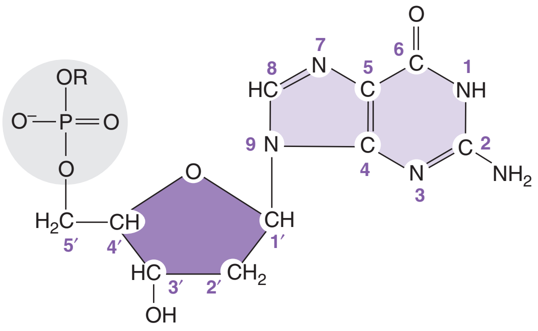

OR
O⁻—P=O
O
H_2C—CH
5′
4′
HC
3′
OH
2′
CH_2
CH
1′
O
9 N
8 HC
7 N
5 C
4 C
6 C
O
1 NH
2 C
NH_2
3 N

FIGURE 1.4 Carbon position numbering of a nucleotide monophosphate. The base carbons are numbered 1 through 9. The sugar carbons are numbered 1′ to 5′. The phosphate group on the 5′ carbon and the hydroxyl group on the 3′ carbon form phosphodiester bonds between bases. The 1′ carbon holds the nitrogen base.

is a left-handed helix with 12 bp per turn and altered geometry of the sugar-base bonds. Z-DNA has been observed in areas of chromosomes where the DNA is under torsional stress from unwinding for transcription or other metabolic functions.

Watson–Crick base pairing (purine:pyrimidine hydrogen bonding) is not limited to the ribofuranosyl nucleic acids, those found in our genetic system. Natural nucleic acid alternatives can also display the basic chemical properties of DNA (and RNA). Theoretical studies have addressed such chemical alternatives to DNA and RNA components. An example is the pentopyranosyl-(2′→4′) oligonucleotide system that exhibits stronger and more selective base pairing than DNA or RNA.[5] The study of nucleic acid alternatives has practical applications. For example, protein nucleic acids, which have a carbon-nitrogen peptide backbone replacing the sugar-phosphate backbone,[6,7] can be used in the laboratory as alternatives to DNA and RNA hybridization probes.[8,9] They have also been proposed as potential enzyme-resistant alternatives to RNA in antisense RNA therapies.[10]

Two DNA chains form **hydrogen bonds** with each other in a specific way. Guanines in one chain form three hydrogen bonds with cytosines in the opposite chain, and adenines form two hydrogen bonds with thymines (see Fig. 1.1B). In this way, single nucleic acid strands can bind or hybridize to single strands that have the corresponding bases. Hydrogen bonds between nucleotides are the key to the specificity of most nucleic acid–based tests used in the molecular laboratory. Specific hydrogen bond formation is also how the information held in the linear order of the nucleotides is maintained. As DNA is polymerized, each nucleotide to be added to the new DNA strand hydrogen bonds with the **complementary** nucleotide on the parental strand (A:T, G:C). In this way the parental DNA strand can be replicated without loss of the nucleotide order. Base pairs other than A:T and G:C or mismatches (e.g., A:C, G:T, A:A) can distort the DNA helix and disrupt the maintenance of sequence information.

Advanced Concepts

In addition to the four commonly occurring nucleotide bases, modified bases are also often found in nature. Base modifications have significant effects on phenotype. Some modified bases result from damage to DNA; others are naturally modified for specific functions or to affect gene expression, as will be discussed in later sections.

Modified nucleotides are used by bacteria and viruses as a primitive immune system that allows them to distinguish their own DNA from that of the host or invaders (restriction-modification [RM] system). Recognizing its own modifications, the host can target unmodified DNA for degradation.

Due to their effects on enzymes that metabolize DNA, modified nucleosides have been used effectively for clinical applications (Fig. 1.5). The anticancer drugs 5-bromouridine (5BrdU) and cytosine arabinoside (cytarabine, ara-C) are modified thymidine and cytosine

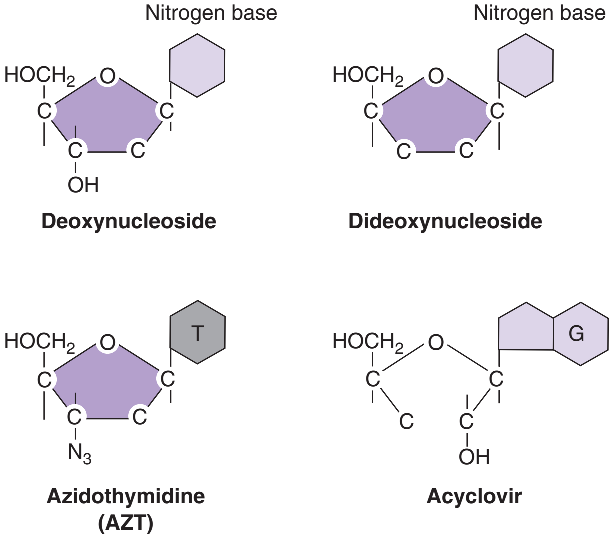


FIGURE 1.5 Substituted nucleosides used in the clinic and the laboratory. Dideoxynucleosides are used as laboratory reagents. Azidothymidine is an antiviral drug that inhibits the human immunodeficiency virus and is used to treat AIDS. Another antiviral, acyclovir, inhibits the growth of herpes viruses.

nucleosides, respectively. Azidothymidine (Retrovir, AZT), cytosine, 2′,3′-dideoxy-2′-fluoro (ddC), and 2′,3′-dideoxyinosine (Videx, ddI), drugs used to treat patients with human immunodeficiency virus (HIV) infections, are modifications of thymidine and cytosine and a precursor of adenine, respectively. An analog of guanosine, 2-amino-1,9-dihydro-9-[(2-hydroxyethoxy)methyl]-6*H*-purin-6-one (Acyclovir, Zovirax) is a drug used to combat herpes simplex virus and varicella-zoster virus.

In the laboratory, nucleosides can be modified for the purposes of labeling or detecting DNA molecules, sequencing, and other applications. The techniques used for these procedures will be discussed in later chapters.

Nucleic Acid

Nucleic acid is a macromolecule made of nucleotides bound together by the phosphate and hydroxyl groups on their sugars. A nucleic acid chain grows by the attachment of the 5′ phosphate group of an incoming nucleotide to the 3′ hydroxyl group of the last nucleotide on the growing chain (Fig. 1.6). Addition of nucleotides in this way gives the DNA chain a **polarity;** that is, it has a 5′ phosphate end and a 3′ hydroxyl end. We refer to DNA as oriented in a 5′ to 3′ direction, and the linear sequence of the nucleotides, by convention, is read in that order.

Advanced Concepts

The sugar-phosphate backbones are arranged at specific distances from one another in the double helix (see Fig. 1.1). The two regions formed in the helix by the backbones are called the **major groove** and **minor groove.** The major and minor grooves are sites of interaction with the many proteins that bind to specific nucleotide sequences in DNA (binding or **recognition sites**). The double helix can also be penetrated by intercalating agents, molecules that slide transversely into the center of the helix. **Denaturing agents** such as formamide and urea displace the hydrogen bonds and separate the two strands of the helix.

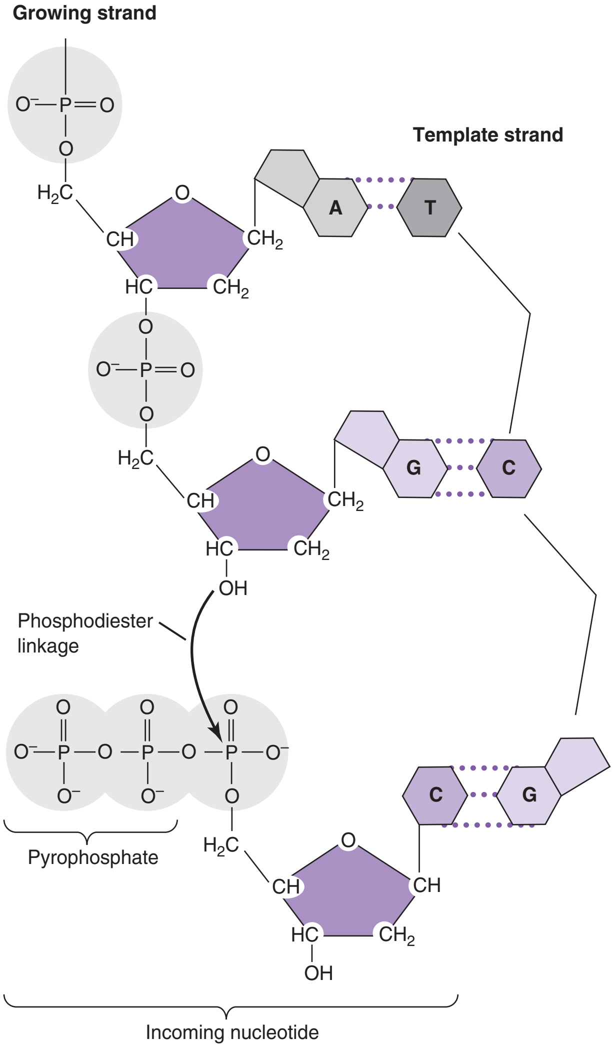


FIGURE 1.6 DNA replication is a template-guided polymerization catalyzed by DNA polymerase. The new strand is synthesized in the 5′ to 3′ direction, reading the template strand in the 3′ to 5′ direction.

DNA found in nature is mostly double stranded. Two strands exist in opposite 5′ to 3′/3′ to 5′ orientation, held together by the hydrogen bonds between their respective bases (A with T and G with C). The bases are positioned such that the sugar-phosphate chain that connects them **(sugar-phosphate backbone)** is oriented in a spiral or helix around the nitrogen bases (see Fig. 1.1).

The DNA double helix represents two versions of the information stored in the form of the order or sequence of the nucleotides on each chain. The sequences of the two strands that form the double helix are complementary, not identical (Fig. 1.7). They are in **antiparallel** orientation, with the 5′ end of one strand at the 3′ end of the other (Fig. 1.8). The formation of hydrogen bonds between two complementary strands of DNA is called **hybridization.** Single strands of DNA with identical sequences will not hybridize with each other. Later sections will describe the importance of this when designing assays.

PO 5′ GTAGCTCGCTGAT 3′ OH
HO 3′ CATCGAGCGACTA 5′ OP

FIGURE 1.7 Homologous sequences are not identical and are oriented in opposite directions.

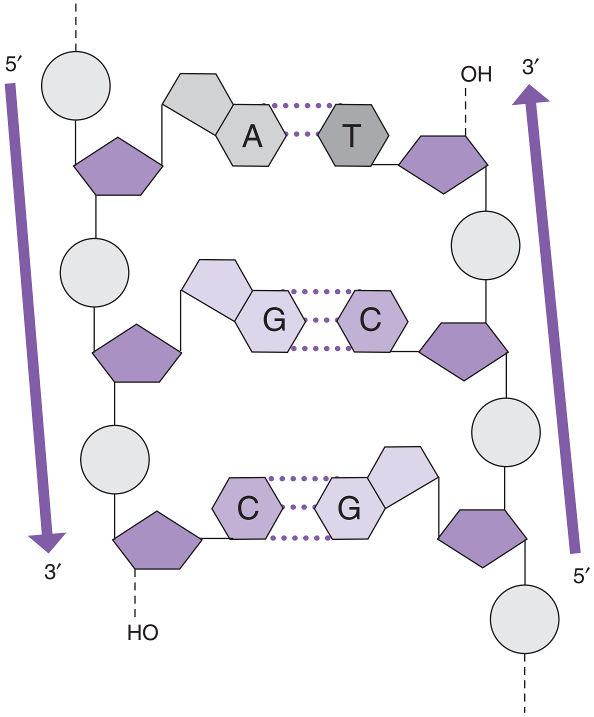


FIGURE 1.8 Because DNA synthesis proceeds from the 5′ phosphate group to the 3′ hydroxyl group and the template strand is copied in the opposite (3′ to 5′) direction, the new double helix consists of the template strand and the new daughter strand oriented in opposite directions from one another.

DNA REPLICATION

The two DNA strands of a double helix have an antiparallel orientation because of the way DNA is replicated. As DNA synthesis proceeds in the 5′ to 3′ direction, DNA polymerase, the enzyme responsible for polymerizing the nucleotide chains, uses a guide, or **template,** to determine which nucleotides to add to the chain. The enzyme reads the template in the 3′ to 5′ direction. The resulting double strand, then, will have a parent strand in one orientation and a newly synthesized strand arranged in the opposite orientation.

As Watson and Crick predicted, **semi-conservative** replication is the key to maintaining the sequence of the nucleotides in DNA through new generations. It is important that this information, in the form of the DNA sequence, be transferred faithfully at cell division. The replication apparatus is designed to copy the DNA strands in an orderly way with minimal errors before each cell division.

Historical Highlights

Before the double helix was determined, Erwin Chargaff[11] made the observation that the amount of adenine in DNA corresponded to the amount of thymine and the amount of cytosine to the amount of guanine. Upon the description of the double helix, Watson proposed that the steps in the ladder of the double helix were pairs of bases, thymine with adenine and guanine with cytosine. Watson and Crick, upon publication of their work, suggested that this arrangement was the basis for a copying mechanism. The complementary strands could separate and serve as guides or templates for producing complementary strands.

In the process of replication, DNA is first unwound from the helical duplex so that each single strand may serve as a template for the addition of nucleotides to the new strand (see Fig. 1.6). The new strand is elongated by hydrogen bonding of the complementary incoming nucleotide to the nitrogen base on the template strand and then a nucleophilic attack of the deoxyribose 3′ hydroxyl oxygen on a phosphorous atom of the phosphate group on the hydrogen-bonded nucleotide triphosphate. Orthophosphate is released with the formation of a phosphodiester bond between the new nucleotide and the last nucleotide of the growing chain. The duplicated helix will ultimately consist of one template strand and one newly synthesized strand.

Historical Highlights

A few years after the solution of the double helix, the mechanism of semiconservative replication was demonstrated by Matthew Meselson and Franklin Stahl,[12] using the technique of equilibrium density centrifugation on a cesium gradient. They prepared "heavy" DNA by growing bacteria in a medium containing the nitrogen isotope ^{15}N. After shifting the bacteria into a medium of normal nitrogen (^{14}N), they could separate hybrid ^{14}N:^{15}N DNA molecules synthesized as the bacteria replicated. These molecules were of a specific intermediate density to the ones from bacteria grown only in ^{14}N or ^{15}N. They could differentiate true semiconservative replication from **dispersive** replication by demonstrating that approximately half of the DNA double helices from the next generation grown in ^{14}N were ^{14}N:^{15}N, and half were ^{14}N:^{14}N.

DNA replication proceeds through the DNA duplex with both strands of DNA replicating in a single pass. DNA undergoing active replication can be observed by electron microscopy as a forked structure, or replication fork. Note, however, that the antiparallel nature of duplex DNA and the requirement for the DNA synthesis apparatus to read the template strand in a 3′ to 5′ direction are not consistent with copying of both strands simultaneously in the same direction. The question arises as to how one of the strands of the duplex can be copied in the same direction as its complementary strand that runs antiparallel to it. This problem was addressed in 1968 by Okazaki and Okazaki[13] when studying DNA replication in *Escherichia coli*. In their experiments, small pieces of DNA, about 1,000 bases in length, could be observed by density-gradient centrifugation in actively replicating DNA. The fragments changed into larger pieces with time, showing that they were covalently linked together shortly after synthesis. These small fragments, or **Okazaki fragments,** were the key to explaining how both strands were copied at the replication fork. The two strands of the parent helix are not copied in the same way. While DNA replication proceeds in a continuous manner on the 3′ to 5′ strand, or the **leading strand,** the replication apparatus jumps ahead a short distance (~1,000 bases) on the 5′ to 3′ strand and then copies backward toward the replication fork. The 5′ to 3′ strand copied in a discontinuous manner is the **lagging strand**[14] (Fig. 1.9).

Another requirement for DNA synthesis is the availability of a deoxyribose 3′-hydroxyl oxygen for chain growth. This means that DNA cannot be synthesized de novo; a preceding base must be present to provide the hydroxyl group. This base is provided by another enzyme component of the replication apparatus, **primase.**[15]

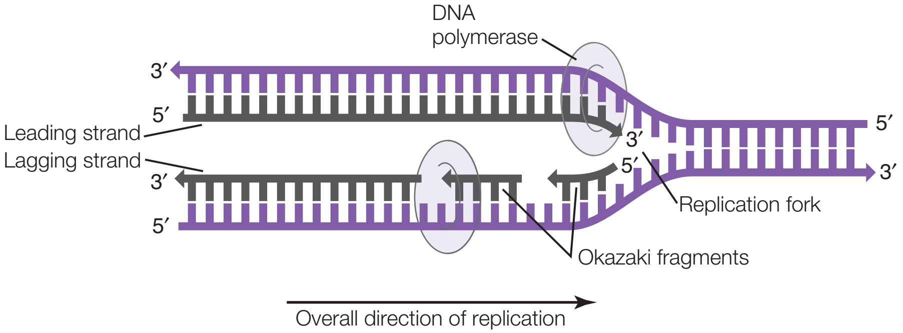


FIGURE 1.9 Simultaneous replication of both strands of the double helix. Both strands are read in the 3′ to 5′ direction. One strand oriented 5′ to 3′ (the lagging strand) is read discontinuously, with the polymerase skipping ahead and reading back toward the replication fork.

Primase is an RNA-synthesizing enzyme that catalyzes the synthesis of short (6 to 11 bp) RNA primers required for priming DNA synthesis. Primase must work repeatedly on the lagging strand to prime synthesis of each Okazaki fragment.

Advanced Concepts

The DNA replication complex (replisome) contains all the necessary proteins for the several activities involved in faithful replication of double-stranded DNA. **Helicase** activity in the replisome unwinds and untangles the DNA for replication. Primase functions either as a separate protein or in a primase-helicase polyprotein in the replisome. The *E. coli* primase, DnaG, transcribes 2,000 to 3,000 RNA primers at a rate of 1 per second in the replication of the *E. coli* genome. Separate polymerase proteins add incoming nucleotides to the growing DNA strands of the replication fork. The details of the synthesis of the lagging strand are not yet clear, although recent evidence suggests that discontinuous replication proceeds by a ratcheting mechanism, with replisome molecules pulling the lagging strand in for priming and copying.[16] Once DNA is primed and synthesized, **RNase H,** an enzyme that hydrolyzes RNA from a complementary DNA strand, removes the primer RNA from the short RNA–DNA hybrid, and the resulting gap is filled by gap-filling DNA polymerase.

Polymerases

The first purified enzyme shown to catalyze DNA replication in prokaryotes was designated DNA polymerase I (pol I). DNA polymerases II (pol II) and III were later characterized, and it was discovered that DNA polymerase III (pol III) was the main polymerizing enzyme during bacterial replication (Table 1.1). The other two polymerases were found to be responsible for the repair of gaps and discontinuities in previously synthesized DNA. It is not surprising that pol I was preferentially purified in those early studies. In in vitro studies where the enzymes were first described, pol II and pol III activity was less than 5% of that of pol I.

TABLE 1.1 Examples of Polymerases Classified by Sequence Homology

Family	Polymerase	Source	Activity
A	Pol I	*E. coli*	Recombination, repair, replication
A	T5 pol, T7 pol	T5, T7 bacteriophage	Replication
A	Pol γ	Mitochondria	Replication
B	Pol II	*E. coli*	Repair
B	Archael	*P. furiosus*	Replication, repair
B	φ29 pol, T4 pol	φ29, T4 bacteriophage	Replication
B	Polα, PolΔ, Polε	Eukaryotes	Repair
B	Viral pols	Various viruses	Repair
C	Pol III core	*E. coli*	Replication
C	dnaE, dnaE^{BS}	*B. subtilis*	Replication
X	Pol β	Eukaryotes	Repair, replication
A	Polη, Polι	Eukaryotes	Bypass replication
Y	Polκ	Eukaryotes	Bypass replication, cohesion
Y	Pol IV, Pol V	*E. coli*	Bypass replication
Y	Rev1, Rad30	*S. cerevisiae*	Bypass replication UV-induced repair
B	Rad 6, Polξ	*S. cerevisiae*	

In vivo, pol III functions as a multi-subunit **holoenzyme.** The holoenzyme works along with a larger assembly of proteins required for priming, initiation, regulation, and termination of the replication process (Fig. 1.10). Two of the 10 subunits of the holoenzyme

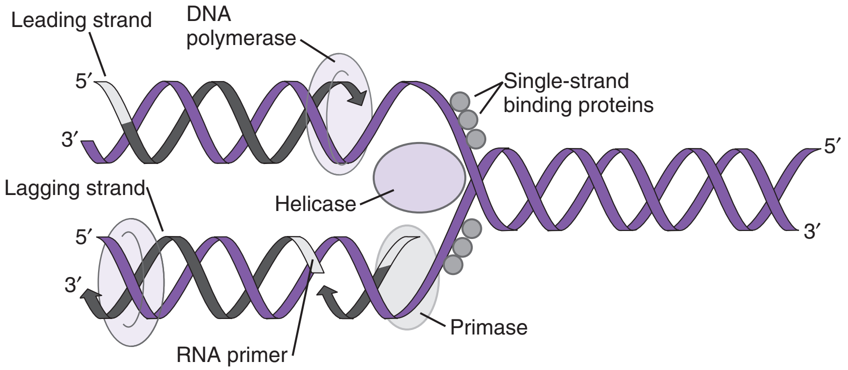


FIGURE 1.10 DNA polymerase activity involves more than one protein molecule. Several cofactors and accessory proteins are required to unwind the template helix, prime synthesis with RNA primers, and protect the lagging strand.

are catalytic DNA polymerizing enzymes, one for the synthesis of the leading strand and one for the synthesis of the lagging strand.[16]

Advanced Concepts

Genome sequencing has revealed that the organization of the proteins in and associated with the holoenzyme is conserved across genera.[17] The conserved nature of the polymerase complex suggests a limited range of possible structures with polymerase activity. It also explains how a bacterial polymerase can replicate DNA from diverse sources. This is important in the laboratory where prokaryote polymerases are used extensively to copy DNA from many different organisms.

Historical Highlights

At a conference on the chemical basis of heredity held at Johns Hopkins University in June 1956, Arthur Kornberg, I. Robert Lehman, and Maurice J. Bessman reported on an extract of *E. coli* that could polymerize nucleotides into DNA in vitro.[14] It was noted that the reaction required preformed DNA and all four nucleotides along with a bacterial protein extract. Any source of preformed DNA would work, bacterial, viral, or animal. At the time, it was difficult to determine whether the new DNA was a copy of the input molecule or an extension of it. During the next 3 years, Julius Adler, Sylvy Kornberg, and Steven B. Zimmerman showed that the new DNA had the same ratio of A-T to G-C bp as the input DNA and was indeed a copy of it. This ratio was not affected by the proportion of free nucleotides added to the initial reaction, confirming that the input or template DNA determined the sequence of the nucleotides on the newly synthesized DNA.

Most DNA polymerases have more than one function, including, in addition to polymerization, pyrophosphorolysis and pyrophosphate exchange, the latter two activities being a reversal of the polymerization process. DNA polymerase enzymes thus have the capacity to synthesize DNA in a 5′ to 3′ direction and degrade DNA in both a 5′ to 3′ and 3′ to 5′ direction (Fig 1.11). The catalytic domain of *E. coli* DNA pol I has two fragments carrying the two functions, a large fragment with the polymerase activity and a small fragment with the exonuclease activity. The large fragment without the exonuclease activity **(Klenow fragment)** has been used extensively in the laboratory for in vitro DNA synthesis.

One purpose of the **exonuclease** function in the various DNA polymerases is to protect the sequence of nucleotides, which must be faithfully copied. Copying errors will result in base changes or **mutations** in the DNA. The 3′ to 5′ exonuclease function is required to ensure that replication begins or continues with a correctly base-paired nucleotide. The enzyme will remove

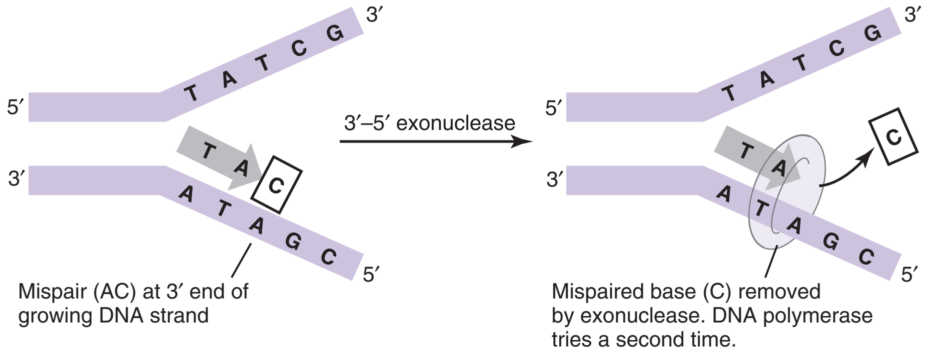

5′
3′
T A T C G
3′
T A
C
A T A G C
5′
3′–5′ exonuclease
Mispair (AC) at 3′ end of growing DNA strand
5′
3′
T A T C G
3′
T A
C
A T A G C
5′
Mispaired base (C) removed by exonuclease. DNA polymerase tries a second time.

FIGURE 1.11 DNA polymerase can remove misincorporated bases during replication using its 3′ to 5′ exonuclease activity.

a mismatch (e.g., A opposite C instead of T on the template) in the primer sequence before beginning polymerization. During DNA synthesis, this exonuclease function gives the enzyme the capacity to **proofread** newly synthesized DNA, that is, to remove a misincorporated nucleotide by breaking the phosphodiester bond and replace it with the correct one.

During DNA replication, *E. coli* DNA pol III can synthesize and degrade DNA simultaneously. At a nick, or discontinuity, in one strand of a DNA duplex, the enzyme can add nucleotides at the 3′ end of the nick while removing nucleotides ahead of it with its 5′ to 3′ exonuclease function (Fig. 1.12). This concurrent synthesis and hydrolysis then moves the nick in one strand of the DNA forward in an activity called **nick translation.** The polymerization and hydrolysis will proceed for a short distance until the polymerase is dislodged. The nick can then be re-closed by DNA **ligase,** an enzyme that forms phosphodiester bonds between existing DNA strands. Nick translation is often used in vitro as a method to introduce labeled nucleotides into DNA molecules. The resulting labeled products are used for DNA detection in hybridization analyses.

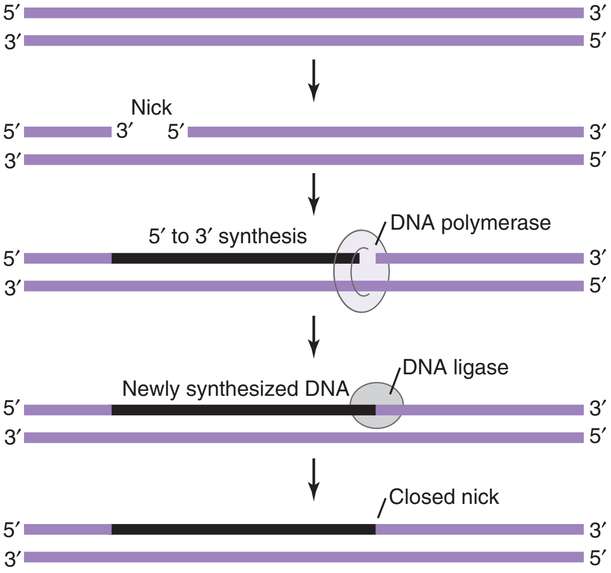

5′
3′
3′
5′
Nick
3′
5′
5′ to 3′ synthesis
DNA polymerase
Newly synthesized DNA
DNA ligase
Closed nick

FIGURE 1.12 Nick translation of DNA. DNA polymerase extends the 3+ end of a nick in double-stranded DNA with a newly synthesized strand (gray) while digesting the original strand from the 5+ end. After polymerization, the nick is closed by DNA ligase.

Advanced Concepts

Like prokaryotes, eukaryotic cells contain multiple polymerase activities. Two polymerase protein complexes, designated α and β, are found in the nucleus and one, γ, in the mitochondria. The three polymerases resemble prokaryotic enzymes, except they have less demonstrable exonuclease activity. A fourth polymerase, δ, originally isolated from bone marrow, has 3′ to 5′ exonuclease activity. Polymerase α, the most active, is identified with chromosome replication, and β and δ are associated with DNA repair.

Advanced Concepts

After replication, distortions in the DNA duplex caused by mismatched or aberrantly modified bases are removed by the 5′ to 3′ exonuclease function of repair polymerases such as DNA pol I. This activity degrades duplex DNA from the 5′ end and can also cleave diester bonds several bases from the end of the chain. It is important for removing lesions in the DNA duplex such as thymine or **pyrimidine dimers,** boxy structures formed between adjacent thymines or cytosines and thymines on the same DNA strand that are induced by exposure of DNA to ultraviolet light. If these structures are not removed, they can disrupt subsequent transcription and replication of the DNA strand.

One type of DNA polymerase, **terminal transferase,** can synthesize polynucleotide chains without a template. This enzyme will add nucleotides to the end of a DNA strand in the absence of hydrogen base pairing with a template. The initial synthesis of a large dA-dT polymer by terminal transferase was a significant event in the history of DNA polymerase studies.[18] Terminal transferase is used in the laboratory to generate 3′-end labeled DNA species.

DNA polymerases play a central role in modern biotechnology. Cloning and some amplification and sequencing technologies require DNA polymerase activity. The prerequisite for specific polymerase characteristics has stimulated the search for new polymerases and the engineering of available polymerase enzymes. Polymerases from various sources are classified into families (A, B, C, X) based on sequence structure.[19,20] A short summary is shown in Table 1.1. Other classifications are based on structural similarities. Polymerases in the A and B family are most useful for biotechnological engineering because the polymerase activity of these enzymes is contained in a **monomeric** or single protein. Chemical manipulation of the amino acid structure produces polymerases with characteristics that are useful in the laboratory. These include altered **processivity** (staying with the template longer to make longer products), **fidelity** (faithful copying of the template), and substrate specificity (affinity for altered nucleotides).[20]

ENZYMES THAT METABOLIZE DNA

Once DNA is polymerized, it is not static. The information stored in the DNA must be tapped selectively to make RNA and, at the same time, protected from damage. Protective systems in prokaryotes eliminate foreign DNA, such as infecting bacteriophages or other viruses. Enzymes that modify and digest foreign DNA prevent infection. In sexually reproducing organisms, the mixing of parental DNA generates genetic diversity **(hybrid vigor)** in the offspring. This process requires cutting and reassembly of the DNA strands in advance of cell division and **gamete** formation. A host of enzymes performs these and other functions during various stages of the cell cycle. Some of these enzymes, including DNA polymerase, have been isolated for in vitro manipulation of DNA in the laboratory. They are key tools of recombinant DNA technology, the basis for commonly used molecular techniques.[21]

Restriction Enzymes

Genetic engineering was stimulated by the discovery of deoxyriboendonucleases, or **endonucleases.** Endonucleases break the sugar-phosphate backbone of DNA.

Advanced Concepts

There are several types of endonucleases. Some prefer single-stranded DNA, and some prefer double-stranded DNA. Repair endonucleases function at areas of distortion in the DNA duplex, such as baseless (apurinic or apyrimidic) sites on the DNA backbone, **thymine dimers,** or mismatched bases. Because the chemical structure of DNA is the same in all organisms, most enzymes are active on DNA from diverse sources.

Restriction enzymes are endonucleases that recognize specific base sequences and break or restrict the DNA polymer at the sugar-phosphate backbone. These

TABLE 1.2 Types of Restriction Enzymes

Type	Methylation Activity	Binding Site Example	Cut Site	Example
I	+	A^{Me}CN^{6}G T	Variable distance from binding	EcoKI, CfrAI
II	–	GAATTC GCCN^{5}GGC	Within binding	EcoR1, BglII
IIS	–	ACCTGC	1–20 bases from binding	BfuAI
IIG	+	CTGAAG CGAN^{6}TGC	To one side or both sides of binding	AcuI BglI
III	+	CGAAT CTGATG (hemimethylated)	25–27 bp 3′ to binding	HinfIII PstIII
IV	+	GGGAC G^{Me}CCGC	Asymetrical	Bsplu11III SauUSI

enzymes were originally isolated from bacteria, where they function as part of a primitive immune system to cleave foreign DNA entering the bacterial cell. The ability of the cell to recognize foreign DNA depends on both DNA sequence recognition and DNA methylation.

Restriction enzymes are named for the organism from which they were isolated. For example, *Bam*HI was isolated from *Bacillus amyloliquefaciens* H, *Hind*III from *Haemophilus influenzae* Rd, *Sma*I from *Serratia marcescens* Sb$_b$, and so forth.

Restriction endonucleases have been classified into four general types (Table 1.2). Type I restriction enzymes have both nuclease and methylase activity in a single enzyme. They bind to host-specific DNA sites of 4 to 6 bp separated by 6 to 8 bp and containing methylated adenines. The site of cleavage of the DNA substrate can be over 1,000 bp from this binding site. An example of a type I enzyme is *EcoK* from *E. coli* K 12. It recognizes the following site:

5′-A^{Me} C N N N N N N G T   G C
T   G N N N N N N C A^{Me} C G-5′

where N represents nonspecific nucleotides, and the adenine residues are methylated (A^{Me}).

Although not completely characterized, type III restriction enzymes resemble type I enzymes in their ability to both methylate and restrict (cut) DNA. Like type I, they have multiple subunits, including helicase (unwinding) activity.[22] Recognition sites for these enzymes are asymmetrical, and the cleavage of the substrate DNA occurs 24 to 26 bp from the site to the 3′ side. An example of a type III enzyme is *Pst*IIII from *P. stuartii*.[22,23] It recognizes the following site:

5′C T G A T   G
G A C T A^{Me} C 5′

where the adenine methylation occurs on only one strand. The enzyme cuts the DNA 25 to 26 bp 3′ to the recognition site.

Type IV restriction enzymes have similar subunit structures and enzyme requirements. Type IV enzymes have cutting and methyltransferase functions. An example of type IV restriction enzymes is *Bse*MII from *Bacillus stearothermophilus*. The *Bse*MII target sequence is

5′C T C A G
G A G T C 5′

The restriction endonuclease function of *Bse*MII cuts one strand of DNA 10 bp following the recognition sequence. *Bse*MII also has a methylation function, adding methyl groups to both of the adenine residues in the target sequence.[24]

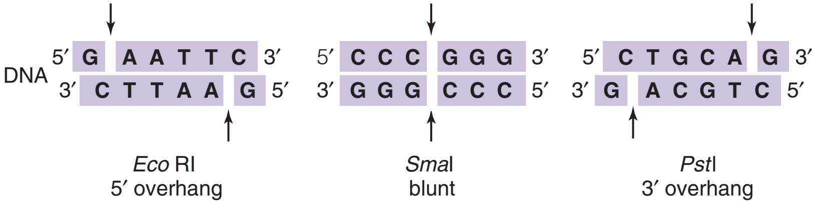


FIGURE 1.13 Type II restriction enzymes recognize symmetrical DNA sequences and cut the sugar-phosphate backbone in different ways, leaving no single strands at the cut site (blunt ends) or 5′ or 3′ overhanging single-stranded ends. Exposed single-stranded ends are "sticky" ends that can hybridize with complementary overhangs.

Type II restriction enzymes are those used most frequently in the laboratory. These enzymes do not have inherent methylation activity. They bind as simple dimers to symmetrical 4- to 8-bp DNA recognition sites. These sites are palindromic in nature; that is, they read the same 5′ to 3′ on both strands of the DNA (Fig. 1.13), referred to as bilateral symmetry. Type II restriction enzymes cleave the DNA directly at the binding site, producing fragments of predictable size.

Type II restriction enzymes have been found in almost all prokaryotes, but none, to date, have been found in eukaryotes. The specificity of their action and the hundreds of enzymes available that recognize numerous sites are key factors in the ability to perform DNA recombination in vitro. Cutting DNA at specific sequences is the basis of many procedures in molecular technology, including mapping, cloning, genetic engineering, and mutation analysis. Restriction enzymes are frequently used in the medical laboratory, for example, in the analysis of gene rearrangements and in mutation detection.

Although all type II restriction enzymes work with bilateral symmetry, their patterns of double-stranded breaks differ (see Fig. 1.13). Some enzymes cut the duplex with a staggered separation at the recognition site, leaving 2- to 4-base single-strand overhangs at the ends of the DNA. The single-strand ends can hybridize with the complementary ends of other DNA fragments, directing the efficient joining of the cut ends. Because of their ability to form hydrogen bonds with complementary overhangs, these cuts are said to produce "sticky ends" at the cut site. Another mode of cutting separates the DNA duplex at the same place on both strands, leaving flush, or blunt, ends. These ends can be rejoined as well, although not as efficiently as sticky ends.

Advanced Concepts

Type II enzymes are subclassified by their characteristic activities. Type IIS restriction endonucleases (shifted cleavage, e.g., *Fok* I, *Alw1*) recognize 5- to 7-bp nonpalindromic sequences and cut DNA within 20 bases of the recognition site. Type IIT enzymes are composed of two subunits, each of which contains one catalytic site. Type IIM restriction enzymes (e.g., *Dpn1*) cut methylated DNA. Methylation-specific enzymes are a useful tool in detecting methylation in DNA.

The advantage of blunt ends for in vitro recombination is that blunt ends formed by different enzymes can be joined, regardless of the recognition site. This is not true for sticky ends, which must have matching overhangs. Sticky ends can be converted to blunt ends using DNA polymerase to extend the recessed strand in a sticky end, using the nucleotides of the overhang as a template or by using a single-strand exonuclease to remove the overhanging nucleotides. Synthetic short DNA fragments with one blunt end and one sticky end (adaptors) can be used to convert blunt ends to specific sticky ends.

Restriction enzymes can be used for mapping a DNA fragment, as will be described in later sections. The collection of fragments generated by digestion of a given DNA fragment (e.g., a region of a human chromosome) with several restriction enzymes will be unique to that DNA. This is the basis for forensic identification and paternity testing using restriction-fragment analysis of human DNA.

Advanced Concepts

About half of all bacteria (and 90% of archeaea) have adaptive immune systems that function with the restriction-modification systems. These **clustered, regularly interspaced, short palindromic repeats (CRISPR)**-Cas protein-based immune systems not only recognize exogenous DNA but also maintain sequence records of previous infections, similar to the function of vertebrate immunity.[25]

CRISPR is mediated by DNA sequence arrays flanked by *Cas* genes, the latter encoding nucleases and helicases. There are six types of CRISPR-Cas structures, classified as class 1 (types I, III, and IV), which have multiple subunit complexes, and class 2 (types II, V, and VI), which have single subunits.[26] Foreign DNA is inserted into the host's genome at these sequence arrays or spacers. Sequences matching the host spacers are **protospacers.** Protospacesrs are recognized by the host through a 2-5 nucleotide **protospacer adjacent motif (PAM)** next to the protospacer sequence.

DNA Ligase

DNA ligase catalyzes the formation of a phosphodiester bond between adjacent 3′-hydroxyl and 5′-phosphoryl nucleotide ends. Its existence was predicted by the observation of replication, recombination, and repair activities in vivo. These operations require reunion of the DNA backbone after discontinuous replication on the lagging strand, strand exchange, or repair synthesis. In 1967, DNA ligase was discovered in five different laboratories.[27] The isolated enzyme was found to catalyze the end-to-end joining of separated strands of DNA.

Other DNA Metabolizing Enzymes

Nucleases

In contrast to endonucleases, exonucleases degrade DNA from free 3′-hydroxyl or 5′-phosphate ends. Consequently, they will not work on closed circular DNA. These enzymes are used, under controlled conditions, to manipulate DNA in vitro,[28] for instance, to make stepwise deletions in linearized DNA or to modify DNA ends after cutting with restriction enzymes. Exonucleases have different substrate requirements and will therefore degrade specific types of DNA ends.

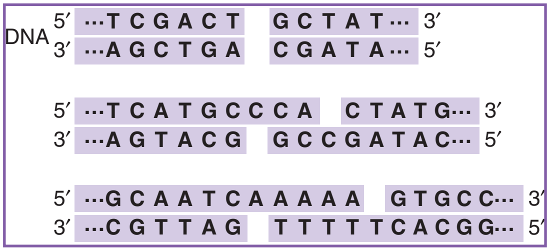
Exonuclease I from *E. coli* degrades single-stranded DNA from the 3′-hydroxyl end into mononucleotides. Its activity is optimal on long single-stranded ends, slowing significantly as it approaches a double-stranded region. It can be used to remove single-stranded excess primers from double-stranded reaction products of DNA copying or amplification procedures.

Exonuclease III from *E. coli* removes 5′ mononucleotides from the 3′ end of double-stranded DNA in the presence of Mg^{2+} and Mn^{2+}. It also has some endonuclease activity, cutting DNA at apurinic sites. Exo III removes nucleotides from blunt ends, recessed ends, and nicks but will not digest 3′ overhangs. Exo III was used in the research setting to create nested deletions in double-stranded DNA or to produce single-stranded DNA for dideoxy sequencing. Exo III can also promote site-specific mutagenesis.

Exonuclease VII from *E. coli* digests single-stranded DNA from either the 5′-phosphate or 3′-hydroxyl end. It is one of the few enzymes with 5′ exonuclease activity. Exo VII can be employed to remove long single strands protruding from double-stranded DNA or single strands from mixtures of single- and double-stranded DNA.

Historical Highlights

The initial analysis of the reaction joining separate DNA helices was performed with physically fractured DNA with no complementary bases at the ends of the fragments. The joining reaction required the chance positioning of two adjacent ends and was, therefore, not very efficient. A better substrate for the enzyme would be ends that could be held together before ligation (i.e., by hydrogen bonds between single strands). H. Gobind Khorana showed that short synthetic segments of DNA with single-strand complementary overhangs joined into larger fragments efficiently.[29] Several investigators noted the increased efficiency of joining of ends of DNA molecules from certain bacterial viruses. These ends have naturally occurring single-stranded overhangs. It was also observed that treatment of DNA ends with terminal transferase to add short runs of A's to one fragment and T's to another increased the efficiency of joining the ends of any two treated fragments. Although not yet available when ligase activity was being studied, the single-strand overhangs left by some restriction enzymes were better substrates for DNA ligase than the blunt ends due to hydrogen bonding of the complementary single-stranded bases.


DNA
5′ ···TCGACT GCTAT··· 3′
3′ ···AGCTGA CGATA··· 5′
5′ ···TCATGCCCA CTATG··· 3′
3′ ···AGTACG GCCGATAC··· 5′
5′ ···GCAATCAAAAA GTGCC··· 3′
3′ ···CGTTAG TTTTTCACGG··· 5′

Substrates for DNA ligase are broken double helices. Blunt ends (top) or noncomplementary overhangs (center) are joined less efficiently than complementary overhangs (bottom). Also note the complementary overhangs in Figure 1.13.

Nuclease Bal31 from *Alteromonas espejiani* can degrade single- and double-stranded DNA from both ends. It can also act as an endonuclease on single-stranded DNA. Because its activity at 20°C is slow enough to control with good resolution, it was useful in research applications to make nested deletions in DNA.

Mung bean nuclease from Mung bean sprouts digests single-stranded DNA and RNA. Because it leaves double-stranded regions intact, it was used to remove overhangs from restriction fragments to produce blunt ends for cloning. Its activity on RNA was used for transcriptional mapping and to resolve hairpins (folds) in RNA.

S1 nuclease from *Aspergillus oryzae* and certain *Neurospora* species is another single-strand–specific nuclease. It hydrolyzes single-stranded DNA or RNA into 5′ mononucleotides. It also has endonuclease capability to hydrolyze single-stranded regions such as gaps and loops in duplex DNA. It was used in early nuclease mapping assays of DNA- and RNA-binding proteins.[30]

recBC nuclease (Exonuclease V) from *E. coli* is an ATP-dependent single- and double-stranded DNA nuclease. Although it has no activity at nicks (short single-strand gaps) in the DNA, it digests single-stranded DNA from either the 3′-hydroxyl or the 5′-phosphate end. It has some endonuclease activity on duplex DNA, generating short fragments, or **oligonucleotides**. At high levels of ATP, this enzyme also unwinds the DNA double helix.

Micrococcal nuclease digests single- and double-stranded DNA and RNA at AT- or AU-rich regions. Although this enzyme can digest duplex DNA, it prefers single-stranded substrates. It is used in the laboratory to remove nucleic acid from crude extracts and also for the analysis of chromatin structure.[31]

Deoxyribonuclease I (DNAse I) from a bovine pancreas digests single- and double-stranded DNA at pyrimidines to oligonucleotides; so, technically, it is an endonuclease. It is used in both research and clinical laboratories to remove DNA from RNA preparations. DNAse I has also been used to detect exposed regions of DNA in DNA protein-binding experiments.

DNA pol I from *E. coli* has exonuclease activity. Formerly called exonuclease II, this activity is responsible for the proofreading function of the polymerase.

Because nucleases are natural components of cellular lysates, it is important to eliminate or inactivate them when preparing nucleic acid specimens for clinical analysis. Most DNA isolation procedures are designed to minimize both endonuclease and exonuclease activity during DNA isolation. Purified DNA is often stored in TE buffer (10 mM Tris-HCl, pH 8.0, 0.1 mM EDTA) to chelate cations required by nucleases for activity.

Helicases

DNA in bacteria and eukaryotes does not exist as the relaxed double helix as shown in Figure 1.1 but as a series of highly organized loops and coils. The release of DNA for transcription, replication, and recombination without tangling is brought about through cutting and re-closing of the DNA sugar-phosphate backbone. These functions are carried out by a series of enzymes called helicases.

As described with restriction endonucleases, the DNA double helix can be broken apart by the separation of the sugar-phosphate backbones in both strands, a double-strand break. When only one backbone is broken (a single-strand break or nick), the broken ends are free to rotate around the intact strand. These ends can be digested by exonuclease activity or extended using the intact strand as a template (nick translation). The nicking and re-closing of DNA by helicases relieve topological stress in highly compacted, or supercoiled, DNA as required, for example, in advance of DNA replication or transcription. Helicases are of two types: **topoisomerases** and gyrases. Topoisomerases interconvert topological isomers or relax supertwisted DNA. Gyrases (type II topoisomerases) untangle DNA through double-strand breaks. They also separate linked rings of DNA (concatemers).

Topoisomerases in eukaryotes have activity similar to that in bacteria but with different mechanisms of cutting and binding to the released ends of the DNA. Because of their importance in cell replication, topoisomerases are the targets for several anticancer drugs, such as camptothecin, the epipodophyllotoxins VP-16 and VM-26, amsacrine, and the intercalating anthracycline derivatives doxorubicin and mitoxantrone.[32] These topoisomerase inhibitors bring about cell death by interfering with the breaking and joining activities of the enzymes, in some cases trapping unfinished and broken intermediates.

Methyltransferases

DNA methyltransferases catalyze the addition of methyl groups to nitrogen bases, usually adenines and cytosines in DNA strands. Most prokaryotic DNA is methylated, or hemimethylated (methylated on one strand of the double helix and not the other), as a means to differentiate host DNA from non-host and to provide resistance to restriction enzymes. Unlike prokaryotic DNA, eukaryotic DNA is methylated in specific regions. In eukaryotes, DNA-binding proteins may limit accessibility or guide methyltransferases to specific regions of the DNA.

Historical Highlights

Enzymatic interconversion of DNA forms was first studied in vitro by observing the action of two *E. coli* enzymes, topoisomerase I[33] and gyrase,[34] on circular plasmids.

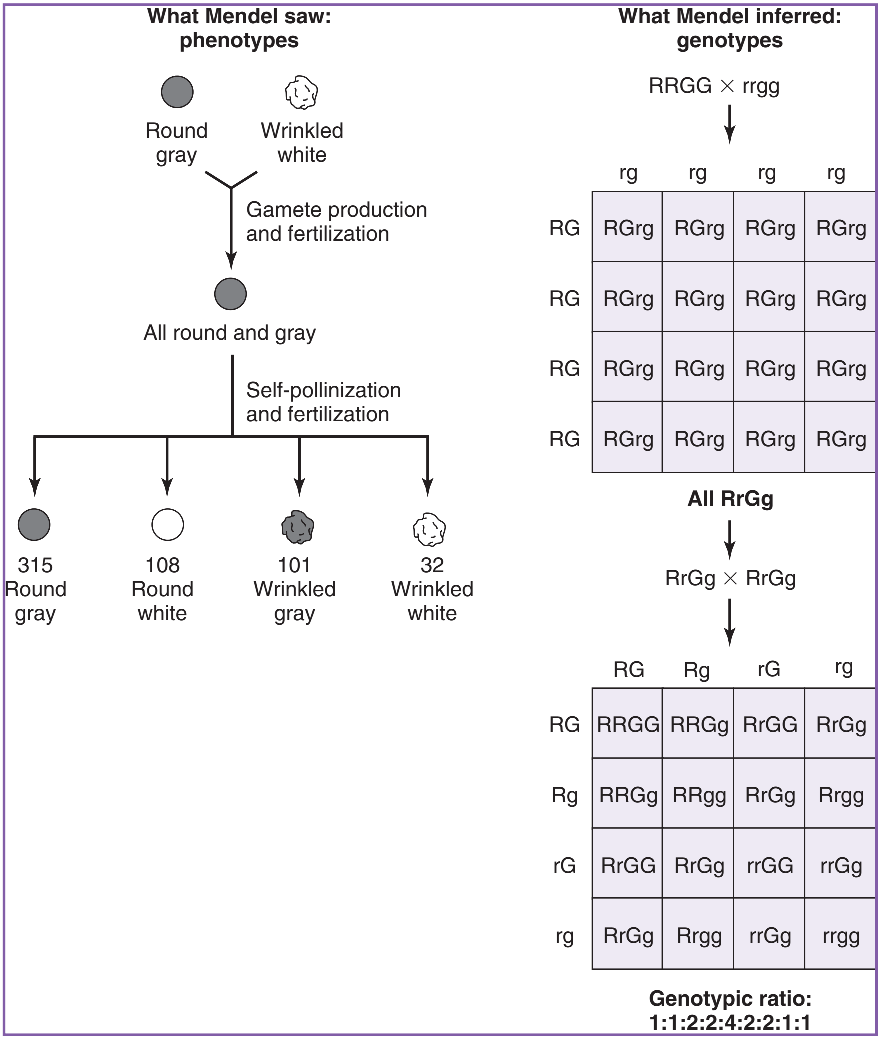
RECOMBINATION IN SEXUALLY REPRODUCING ORGANISMS

Recombination is the mixture and assembly of new genetic combinations. Recombination occurs through the molecular process of crossing over or physical exchange between molecules. A **recombinant** molecule or organism is one that holds a new combination of DNA sequences.

Based on Mendel's laws, each generation of sexually reproducing organisms is a new combination of the parental genomes. The mixing of genes generates genetic diversity, increasing the opportunity for more robust and well-adapted offspring. The beneficial effect of natural recombination is observed in the heterosis, or hybrid vigor, observed in genetically mixed or hybrid individuals compared with purebred organisms.

Historical Highlights

Early studies of recombination were done with whole organisms. Mendel's analysis of peas (*Pisum* species)[35] established the general rules of recombination in sexually reproducing organisms. Mendel could infer the molecular exchange events that occurred in the plants by observing the phenotype of progeny. These observations had been made before, with quantitative predictions of the probability of phenotypes. Mendel proposed that traits are inherited in a particulate manner, rather than blending as was previously thought.


What Mendel saw: phenotypes
Round gray
Wrinkled white
Gamete production and fertilization
All round and gray
Self-pollinization and fertilization
315 Round gray
108 Round white
101 Wrinkled gray
32 Wrinkled white
What Mendel inferred: genotypes
RRGG × rrgg
All RrGg
RrGg × RrGg
Genotypic ratio: 1:1:2:2:4:2:2:1:1

	rg	rg	rg	rg
RG	RGrg	RGrg	RGrg	RGrg
RG	RGrg	RGrg	RGrg	RGrg
RG	RGrg	RGrg	RGrg	RGrg
RG	RGrg	RGrg	RGrg	RGrg

	RG	Rg	rG	rg
RG	RRGG	RRGg	RrGG	RrGg
Rg	RRGg	RRgg	RrGg	Rrgg
rG	RrGG	RrGg	rrGG	rrGg
rg	RrGg	Rrgg	rrGg	rrgg

Mendelian genetics showed that traits are inherited as unit characteristics. The probability of inheriting a given trait can then be calculated from the traits of the parents. *R,* round; *r,* wrinkled; *G,* gray; *g,* white.

Sexually reproducing organisms mix genes in three ways. First, at the beginning of **meiosis,** duplicated chromosomes line up and recombine by crossing over or breakage and reunion of the four DNA duplexes (Fig. 1.14). This generates newly recombined duplexes with genes from each duplicate. Then, the recombined duplexes are randomly assorted into gametes (Fig. 1.15) so that each gamete contains one set of each of the recombined parental chromosomes. Finally, the gamete will merge with a gamete from the other parent carrying

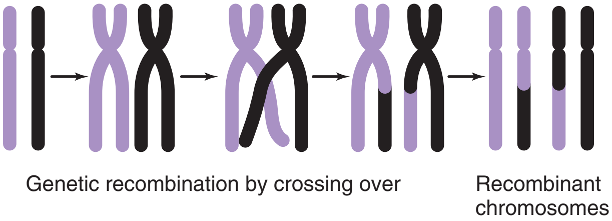

Genetic recombination by crossing over
Recombinant chromosomes

FIGURE 1.14 Generation of genetic diversity by crossing over of homologous chromosomes.

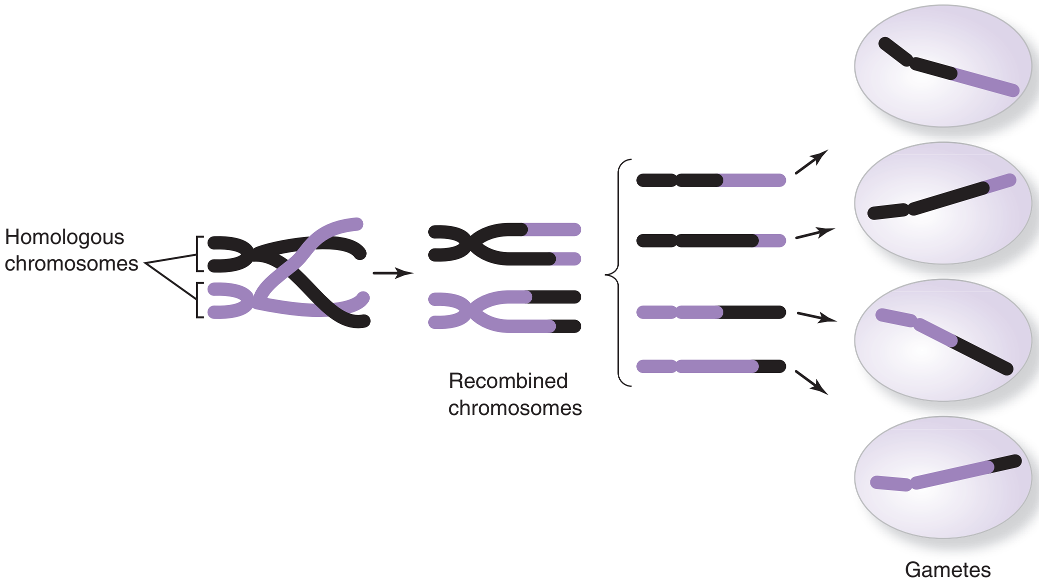


FIGURE 1.15 Recombined chromosomes are randomly assorted into gametes. Twenty-two other chromosomes will be randomly assorted into the four gametes, giving each one a new collection of recombined chromosomes.

its own set of recombined chromosomes. The resulting offspring will contain a new set or recombination of the genes of both parents. The nature of this recombination is manifested in the combinations of inherited traits of subsequent generations.

Recombinant DNA technology is a controlled mixing of genes. Rather than relying on natural mixing of whole genomes, single genes can be altered, replaced, deleted, or moved into new genomes. This directed diversity can produce organisms with predictable traits, as natural purebreds, but with single-gene differences. The ability to manipulate single traits has implications not only in the laboratory but also potentially in the treatment and prevention of disease, for example, through gene therapy.

RECOMBINATION IN ASEXUAL REPRODUCTION

Movement and manipulation of genes in the laboratory began with the study of natural recombination in asexually reproducing bacteria. Genetic information in asexually reproducing organisms can be recombined in three ways: conjugation, transduction, and transformation (Fig. 1.16).

Conjugation

Bacteria that participate in **conjugation** are of two types, or sexes, termed F+ and F–. For conjugation to occur, F– and F+ cells must be in contact with each other. The requirement for contact can be demonstrated by physically separating F+ and F– cells. If this is done, mating does not occur (Fig. 1.17). Microscopically, a filamentous bridge is observed between mating bacteria. Work by J. Lederberg and William Hayes demonstrated polarity in the conjugation process; that is, genetic information could move from F+ to F– bacteria but not from F– to F+ bacteria. The explanation for this was soon discovered. The F+ bacteria had a “fertility factor” that not only carried the information from one cell to another but also was responsible for establishing the physical connection between the mating bacteria. The fertility factor was transferred from F+ to F– bacteria in the mating process so that afterward, the F– bacteria became F+ (Fig. 1.18).

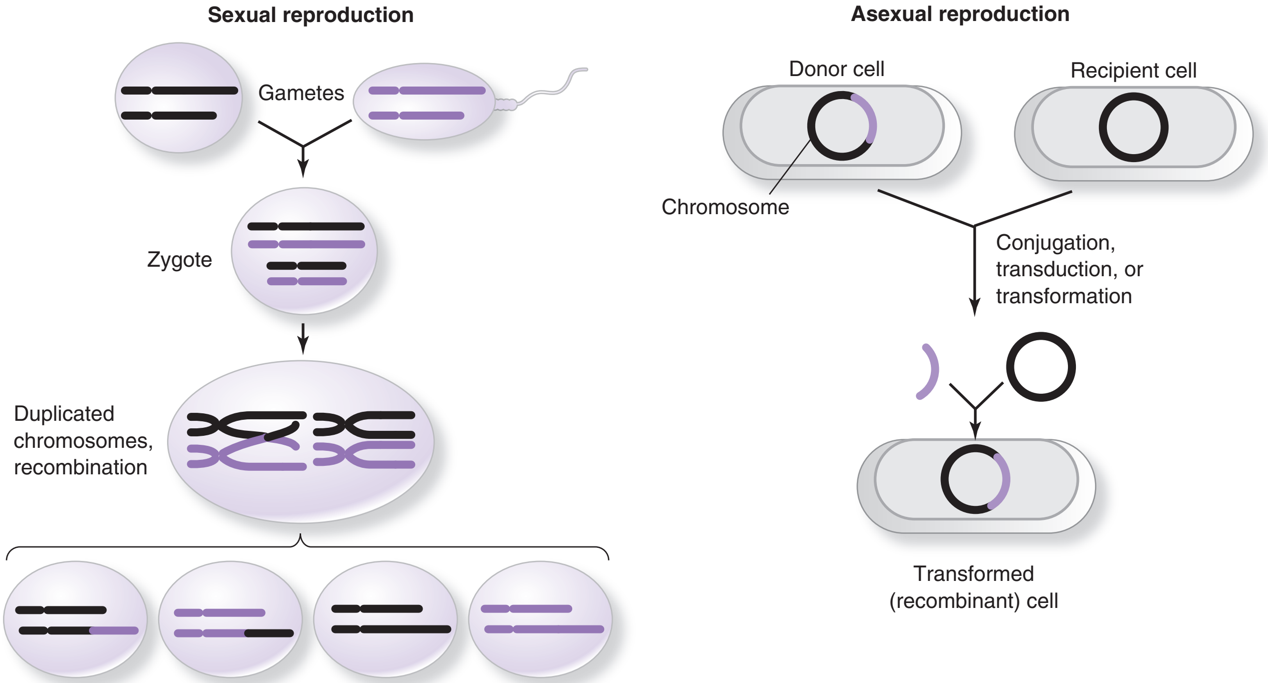

Sexual reproduction
Gametes
Zygote
Duplicated chromosomes, recombination
Asexual reproduction
Donor cell
Recipient cell
Chromosome
Conjugation, transduction, or transformation
Transformed (recombinant) cell

FIGURE 1.16 Recombination in sexual *(left)* and asexual *(right)* reproduction.

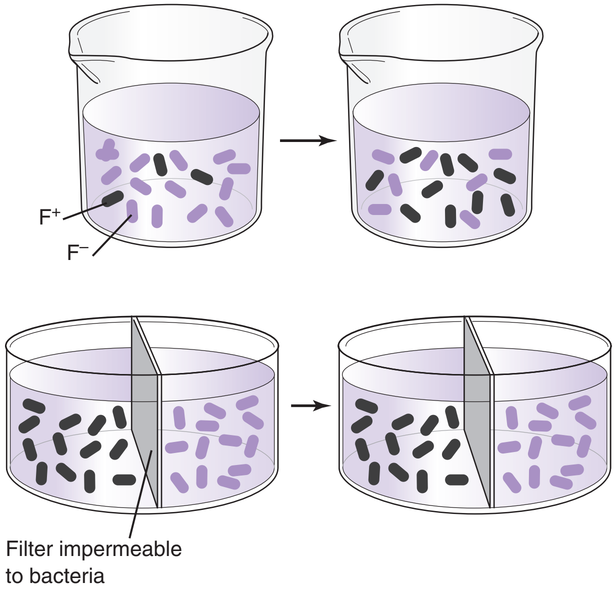

F+
F–
Filter impermeable to bacteria

FIGURE 1.17 Conjugating cells must be in physical contact with each other *(top)* for successful transfer of the F+ phenotype. If cells are separated by a membrane *(bottom)*, F– bacteria do not become F+.

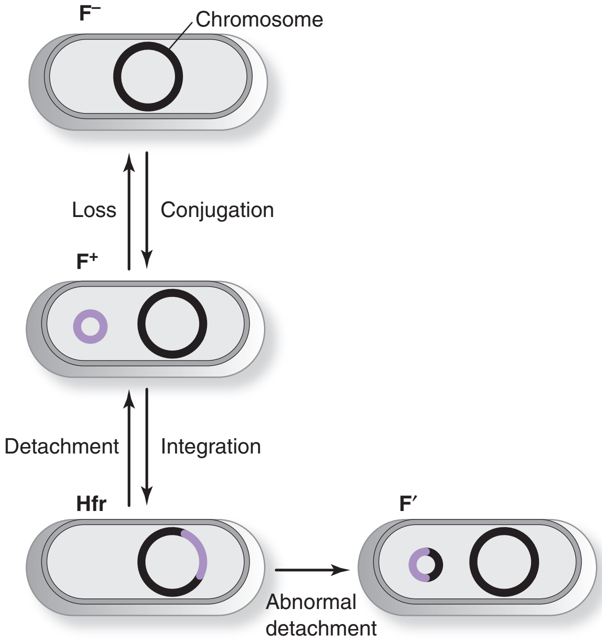
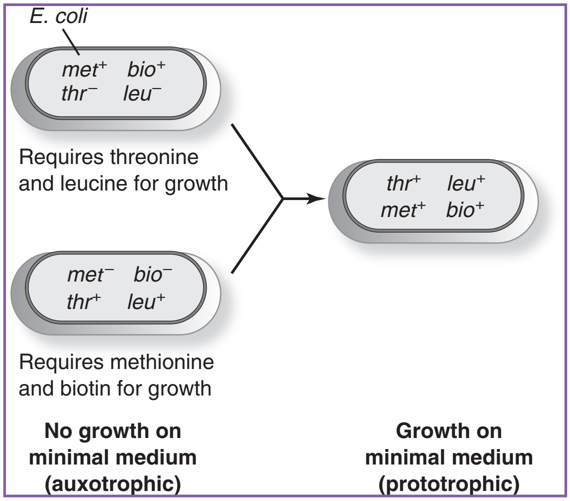

F–
Chromosome
Loss
Conjugation
F+
Detachment
Integration
Hfr
F′
Abnormal detachment

FIGURE 1.18 Fertility (the ability to donate genetic information) is controlled by the F factor. The F factor can exist by itself or be integrated into the host chromosome.

Historical Highlights

Historically, recombination was studied through controlled mating and propagation of organisms. George Beadle[36] and others confirmed the connection between the units of heredity and physical phenotype using molds *(Neurospora crassa),* bacteria, and viruses. Joshua Lederberg and Edward L. Tatum[37] demonstrated that bacteria mate and exchange genetic information to produce recombinant offspring. Lederberg and Tatum proved that genetic exchange between organisms was not restricted to the sexually reproducing molds. These early studies first demonstrated the existence of recombination in E. coli.


E. coli
met^+ bio^+
thr^- leu^-
Requires threonine and leucine for growth
met^- bio^-
thr^+ leu^+
Requires methionine and biotin for growth
thr^+ leu^+
met^+ bio^+
No growth on minimal medium (auxotrophic)
Growth on minimal medium (prototrophic)

Transfer of genetic information by conjugation can be demonstrated using double mutants. Bacteria with double mutations that require exogenous methionine and biotin (*met* and *bio*) or threonine and leucine (*thr* and *leu*) cannot grow on a medium without addition of these nutrients (minimal medium). When these strains are mixed together, however, growth occurs. The resulting bacteria have acquired the normal genes (+) through transfer or conjugation.

The **F factor** was shown to be an extrachromosomal circle of double-stranded DNA or **plasmid** carrying the genes coding for construction of the mating bridge. Genes carried on the F factor are transferred across the bridge and simultaneously replicated so that one copy of the F factor remains in the F+ bacteria, and the other is sent into the F– bacteria. After mating, both bacteria are F+. The F factor may be lost or cured during normal cell division, converting an F+ bacteria to the F– state.

The F factor can also insert itself into the host chromosome through a crossover or recombination event. Embedded in the chromosome, the F factor maintains its ability to direct mating and can carry part or all of the host chromosome with it across the mating bridge into F– bacteria. Strains with chromosomally embedded F factors are called high-frequency recombination (Hfr) bacteria. When the embedded F factor in these rarely occurring strains pulls host chromosomal information into recipient bacteria, another recombination event can insert that information into the recipient chromosome, forming a recombinant or new combination of genes of the Hfr and F– bacteria. Hfr bacteria were used in the first mapping studies.

Transduction

In the early 1960s, Francois Jacob and Elie Wollman[38] studied the transmission of units of heredity carried by viruses from one bacterium to another **(transduction).** Just as animal and plant viruses infect eukaryotic cells, bacterial viruses, or **bacteriophages,** infect bacterial cells. The structure of bacteriophage T4 is one example of the specialized protein coats that enable these viruses to insert their DNA through the cell wall into the bacterial cell (Fig. 1.19). Alfred Hershey and Martha Chase confirmed that the DNA of a bacterial virus was the carrier of its genetic determination in the transduction process.[39] Hershey and Chase used ^{35}S to label the viral protein and ^{32}P to label the viral DNA. The experiment showed that viral protein remained outside of the cell, whereas viral DNA entered the cell. Furthermore, ^{32}P-labeled DNA could be detected in new viruses generated in the transduction process (Fig. 1.20). Methods soon developed using bacteriophages to move genetic information between bacteria by growing the phage on one strain of bacteria and then infecting a second strain with those viruses. Transduction is also useful in determining gene order. Seymour Benzer used transduction of the T4 bacterial virus to fine-map genes.[40]

Transformation

Although conjugation and transduction were the methods for the initial study of the connection between DNA and phenotype, **transformation,** which was first observed in

1928 by Frederick Griffith,[41] is the basis for modern-day recombinant techniques. Griffith was investigating virulence in *Diplococcus* (now known as *Streptococcus*) *pneumoniae*. He had two strains of the bacteria: one with a rough colony type that was avirulent and one with a smooth colony type that was virulent. Griffith intended to use these strains to develop a protective vaccine (Fig. 1.21). He knew that the live smooth-type bacteria were lethal in mice, and the live rough-type were not. If he first killed the smooth-type bacteria by boiling them, virulence was lost, and they were no longer lethal to mice. Surprisingly, when he mixed killed smooth-type and live rough-type bacteria, virulence returned. Furthermore, he could recover live smooth-type bacteria from the dead mice. He concluded that something from the dead smooth-type bacteria had "transformed" the rough-type bacteria into the virulent smooth-type bacteria.

What Griffith had observed was the transfer of DNA from one organism to another without the protection of a conjugative bridge or a viral coat. Fifteen years later, Oswald T. Avery, Colin MacLeod, and M. J. McCarty identified the transforming material as DNA.[42,43] They prepared boiled virulent bacterial cell lysates and sequentially treated them with recently discovered enzymes (Fig. 1.22). **Protease** and **ribonuclease** treatment, which degraded protein and RNA, respectively, did not

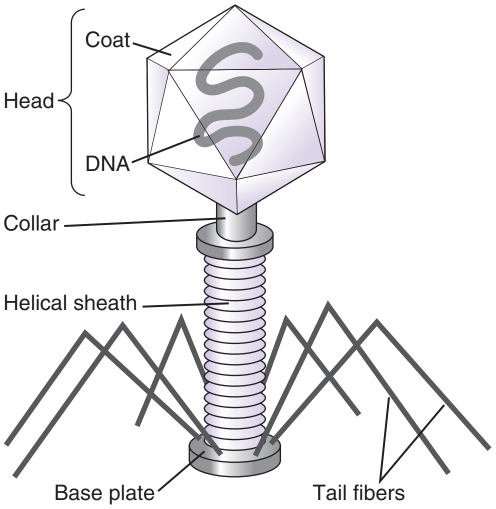

Coat
Head
DNA
Collar
Helical sheath
Base plate
Tail fibers

FIGURE 1.19 Bacteriophage T4 infects specific strains of *E. coli*. It has specialized structures. The tail fibers find the bacterial surface and allow contact of the tail plate and injection of the DNA in the viral head through the sheath into the bacterium.

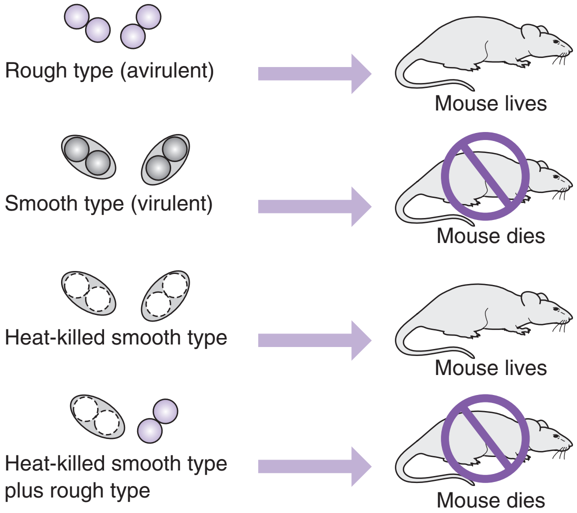

Rough type (avirulent)
Mouse lives
Smooth type (virulent)
Mouse dies
Heat-killed smooth type
Mouse lives
Heat-killed smooth type plus rough type
Mouse dies

FIGURE 1.21 The "transforming factor" discovered by Griffith was responsible for changing the phenotype of the avirulent rough-type bacteria to that of the virulent smooth type.

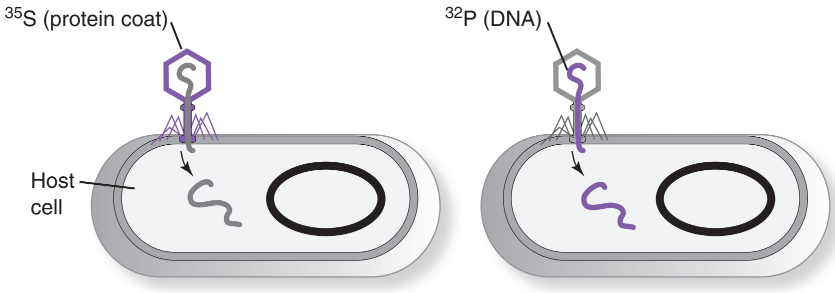

^{35}S (protein coat)
^{32}P (DNA)
Host cell

FIGURE 1.20 Radioactive protein does not enter the host cell during transduction *(left)*. Radioactive DNA, however, does enter and is passed to subsequent generations of viruses.

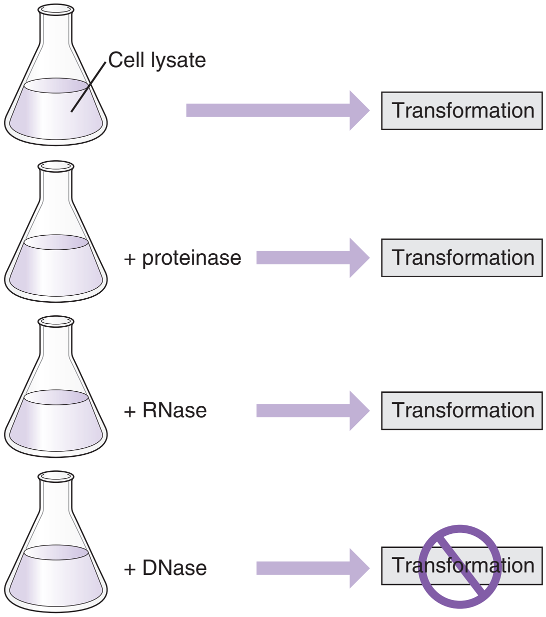

Cell lysate
Transformation
+ proteinase
Transformation
+ RNase
Transformation
+ DNase
Transformation

FIGURE 1.22 Avery, MacLeod, and McCarty showed that destruction of protein or RNA in the cell lysate did not affect the transforming factor. Only destruction of DNA prevented transformation.

affect the transformation phenomenon that Griffith had demonstrated earlier. Treatment with deoxyribonuclease, which degrades DNA, however, prevented transformation. They concluded that the "transforming factor" that Griffith had first proposed was DNA. The transduction experiment of Alfred Hershey and Martha Chase also confirmed their findings that DNA carried genetic traits.

Advanced Concepts

Investigators performing early transformation studies observed the transfer of broken chromosomal DNA from one population of bacterial cells to another. Naked DNA transferred in this way, however, is very inefficient. Unprotected DNA is subject to physical shearing as well as chemical degradation from naturally occurring nucleases, especially on the broken ends of the DNA molecules. Natural transformations are much more efficient because the transforming DNA is in the form of circles or otherwise protected, especially on the ends.

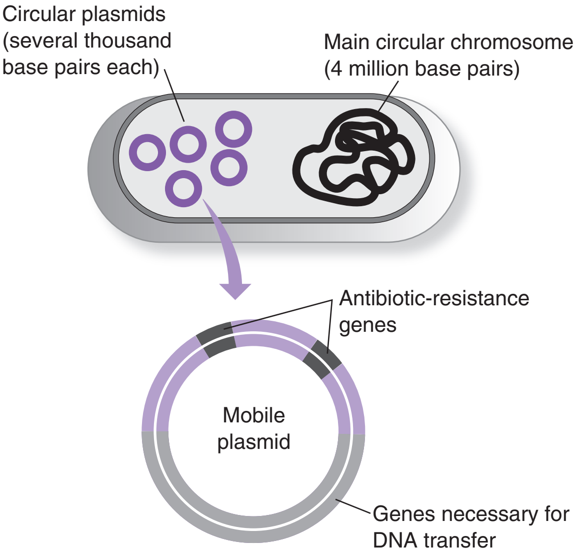

Circular plasmids
(several thousand
base pairs each)
Main circular chromosome
(4 million base pairs)
Antibiotic-resistance
genes
Mobile
plasmid
Genes necessary for
DNA transfer

FIGURE 1.23 Plasmids are small extrachromosomal DNA duplexes that can carry genetic information.

PLASMIDS

DNA helices can assume both linear and circular forms. Most bacterial chromosomes are in circular form, in contrast to chromosomes in higher organisms, such as fungi, plants, and animals, which are mostly linear. A bacterial cell can contain, in addition to its own chromosome complement, extrachromosomal entities, or plasmids (Fig. 1.23). Most plasmids are double-stranded circles of 2,000 to 100,000 bp (2 to 100 kilobase pairs) in size. Plasmids can carry genetic information. Due to their size and effect on the host cell, plasmids carry only a limited amount of information.

The plasmid DNA duplex is compacted, or supercoiled. Breaking one strand of the plasmid duplex, or nicking, will relax the supercoil (Fig. 1.24), whereas breaking both strands will linearize the plasmid. Different physical states of the plasmid DNA can be resolved by distinct **migration** characteristics during gel electrophoresis.

Plasmids were found to be a source of resistant phenotypes in multidrug-resistant bacteria.[44] The demonstration that multiple drug resistance in bacteria can be eliminated by treatment with acridine dyes[45] was the first indication of the **episomal** (plasmid) nature of the

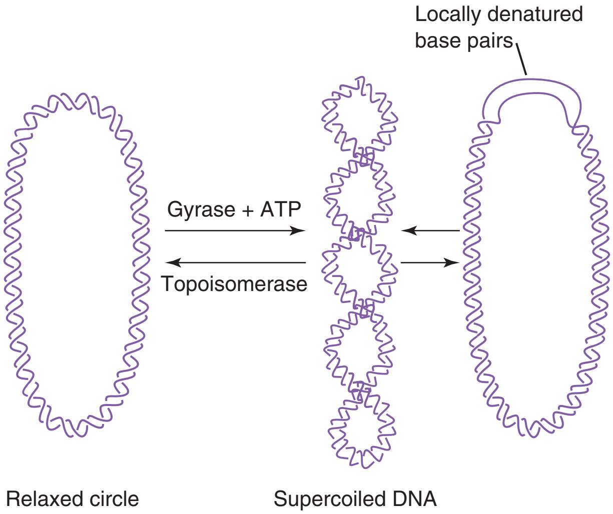


FIGURE 1.24 Supercoiled plasmids can be relaxed by nicking *(left)* or by local unwinding of the double helix *(right)*.

resistance factor, similar to the F factor in conjugation. (Acridine dyes induce the loss of episomes.) In resistant bacteria, plasmids carrying the genes for inactivation or circumvention of antibiotic action were called resistance transfer factors (RTF), or R factors. R factors promote resistance to common antibiotics such as chloramphenicol, tetracycline, ampicillin, and streptomycin. Another class of plasmid, colicinogenic factors, carries resistance to bacteriocins, toxic proteins manufactured by bacteria. The acquisition of the resistance genes from host chromosomes of unknown bacteria is the presumed origin of these resistance factors.[46] Drug-resistance genes are commonly gained and lost from episomes in a bacterial population. Two different R factors in a single cell can undergo a recombination event, producing a new, recombinant plasmid with a new combination of resistance genes.

Plasmids were initially classified into two general types: large plasmids and small plasmids. Large plasmids include the F factor and some of the R plasmids. Large plasmids carry genes for their own transfer and propagation and are self-transmissible. Large plasmids occur in small numbers, one or two copies per chromosome equivalent. Small plasmids are more numerous in the cell, about 20 copies per chromosomal equivalent; however, they do not carry genes directing their maintenance. They rely on high numbers for distribution into daughter cells at cell division or uptake by host cells in transformation.

Advanced Concepts

The circular nature of R factors was demonstrated by buoyant-density centrifugation.[47] Plasmid DNA has a density higher than that of the host chromosome and can be isolated from separate, or satellite, bands in the gradient. Examination of the fractions of the higher-density DNA reveals small circular species. These circles are absent from drug-sensitive bacteria.

Compared with fragments of DNA, plasmids are more efficient vehicles for the transfer of genes from one cell to another. Upon cell lysis, supercoiled plasmids can enter other cells more efficiently. Plasmids are used in recombinant DNA technology to introduce specific traits. By manipulation of the plasmid DNA in vitro, specific genes can be introduced into cells to produce new phenotypes or recombinant organisms. The ability to express genetic traits from plasmids makes it possible to manipulate phenotype in specific ways. As will be described in later chapters, plasmids play a key role in the development of the procedures used in molecular analysis.

Advanced Concepts

Plasmids are found not only in bacteria but in multicellular plants and animals as well. Some viruses, such as the single-stranded DNA virus M13, have a transient plasmid phase in their life cycle. Laboratory techniques requiring single-stranded versions of specific DNA sequences were once based on the manipulation of the plasmid (duplex circle) phase of these viruses and isolation of recombinant single-stranded circles from the virus. This technology was used in methods devised to determine the order or sequence of nucleotides in the DNA chain.

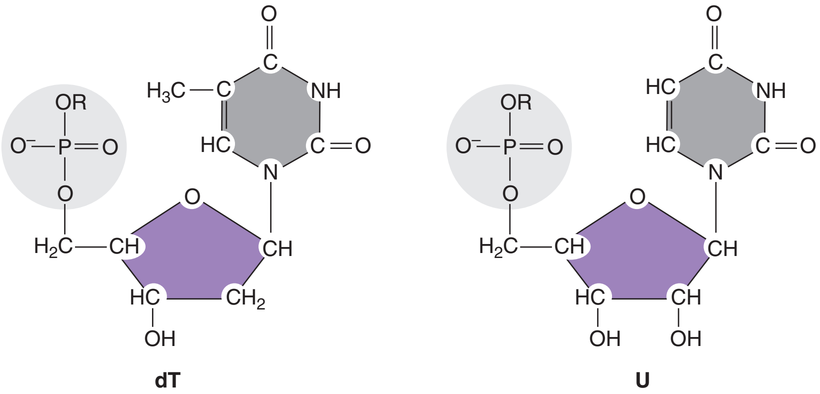


FIGURE 1.25 Uracil (U), the nucleotide base that replaces thymine in RNA, has the pyrimidine ring structure of thymine (dT) minus the methyl group. Uracil forms hydrogen bonds with adenine.

RNA

Ribonucleic acid (RNA) is a polymer of nucleotides similar to DNA. It differs from DNA in the sugar moieties, having **ribose** instead of deoxyribose and, in one nitrogen **base** component, having uracil instead of thymine (thymine is 5-methyl uracil; Fig. 1.25). Furthermore, RNA is synthesized as a single strand rather than as a double helix. Although almost all RNA strands do not have complementary partner strands, they are not completely single stranded. Through internal homologies, RNA species fold and loop upon themselves to take on a double-stranded character that is important for their function. RNA can also pair with complementary single strands of DNA or another RNA and form a double helix.

There are several types of RNAs found in the cell. Ribosomal RNA, messenger RNA, transfer RNA, and small nuclear RNAs have distinct cellular functions. RNA is copied, or transcribed, from DNA.

Advanced Concepts

Evolutionary theory places RNA as the original genetic material from which DNA has evolved. In most organisms, RNA is an intermediate between the storage system of DNA and the proteins responsible for phenotype. One family of RNA viruses, the retroviruses, which include leukemia viruses and the human immunodeficiency virus, have RNA genomes, and in order to replicate using host-cell machinery, they must first make a DNA copy of their genome by reverse transcription.

Transcription

DNA can only store information. In order for this information to be utilized, it must be transcribed and then translated into protein, a process called **gene expression.** A specific type of RNA, messenger RNA (mRNA), carries the information in DNA to the ribosomes, where it is translated into protein.

Transcription is the copying of one strand of DNA into RNA by a process similar to that of DNA replication. This activity, catalyzed by **RNA polymerase,** occurs mostly in interphase. Whereas a single type of RNA polymerase catalyzes the synthesis of all RNA in most prokaryotes, there are three types of RNA polymerases in eukaryotes: RNA polymerase pol I, pol II, and pol III. Pol I and III synthesize noncoding RNA (Table 1.3).[48] Pol II is responsible for the synthesis of **messenger RNA** (mRNA), the type of RNA that carries genetic information to be translated into protein. Evidence suggests that transcription takes place at discrete stations of the nucleus into which the DNA molecules

TABLE 1.3 RNA Polymerases

Enzyme	Template	Product
E. coli RNA polymerase II	DNA	mRNA
RNA polymerase I	DNA	rRNA
RNA polymerase II	DNA	mRNA
RNA polymerase III	DNA	tRNA, snRNA
Mitochondrial RNA polymerase	DNA	mRNA
Mammalian DNA polymerase α	DNA	Primers
HCV RNA polymerase	RNA	Viral genome
Dengue virus RNA polymerase	RNA	Viral genome
PolyA polymerase	None	PolyA tails

TABLE 1.4 Cellular Location and Activity of RNA Pol I, II, and III in Eukaryotes

Type	Location	Products	α-Amanitin
I	Nucleolus	18s, 5.8s, 28s rRNA	Insensitive
II	Nucleus	mRNA, snRNA	Inhibited
III	Nucleus	tRNA, 5s rRNA	Inhibited by high concentration

move (Table 1.4).[2] One of these sites, the nucleolus, is the location of ribosomal RNA synthesis.

Advanced Concepts

DNA must be released locally from histones and the helix unwound in order for transcription to occur. These processes involve the participation of numerous factors, including DNA-binding proteins, transcription factors, histone-modification enzymes, and RNA polymerase.

Transcription Initiation

RNA polymerase and its supporting accessory proteins assemble on DNA at a specific site called the **promoter.** Initiation sites of transcription (RNA synthesis) greatly outnumber DNA initiation sites in both prokaryotes and eukaryotes. There are also many more molecules of RNA polymerase than DNA polymerase in the cell. Although functionally catalyzing the same reaction, RNA polymerases in prokaryotes and eukaryotes differ and work with different supporting proteins to find and bind to DNA in preparation for transcription. In prokaryotes, a **basal transcription complex** comprised of the large and small subunits of RNA polymerase and additional sigma factors assembles at the start site. The eukaryotic transcription complex requires RNA polymerase and up to 20 additional factors for accurate initiation. Initiation of RNA synthesis is regulated in all organisms so that genes are transcribed as required by specific cell types. The regulation of transcription differs in prokaryotes and eukaryotes.

Transcription Elongation

RNA polymerases in both eukaryotes and prokaryotes synthesize RNA using the base sequence of one strand of the double helix as a guide (Fig. 1.26). The complementary strand of the DNA template (that is not copied) has a sequence identical to that of the RNA product (except for the U for T substitution in RNA).

RNA polymerases work more slowly than DNA polymerases (50 to 100 bases/sec for RNA synthesis vs. 1,000 bases/sec for DNA replication) and with less fidelity. The DNA double helix is locally unwound into

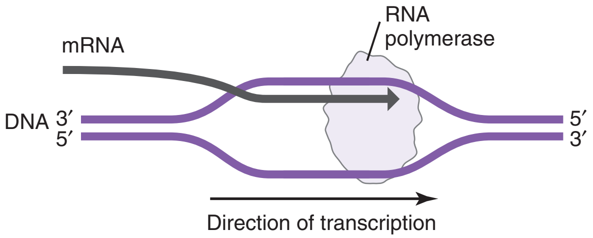


FIGURE 1.26 RNA polymerase uses one strand of the double helix (the antisense strand) as a template for synthesis of RNA. About 10 base pairs of DNA are unwound or opened to allow the polymerase to work.

single strands to allow the assembly and passage of the transcription machinery, forming a transcription bubble.

Unlike DNA synthesis, RNA synthesis does not require priming. The first ribonucleoside triphosphate retains all of its phosphate groups as the RNA is polymerized in the 5′ to 3′ direction. Subsequent ribonucleoside triphosphates retain only the **alpha phosphate,** the one closest to the ribose sugar. The other two phosphate groups are released as orthophosphate during the synthesis reaction.

Transcription Termination

RNA synthesis terminates differently in prokaryotes and eukaryotes. In prokaryotes, some RNA synthesis is responsive to protein products, such that high levels of a gene product induce termination of its own synthesis. Termination is accomplished in some genes by interactions between RNA polymerase and nucleotide signals in the DNA template. In other genes, an additional factor, rho, is required for termination. Rho is a helicase enzyme that associates with RNA polymerase and inactivates the elongation complex at a cytosine-rich termination site in the DNA.[49,50] Rho-independent termination occurs at G:C-rich regions in the DNA, followed by A:T-rich regions. The G:C bases are transcribed into RNA and fold into a short double-stranded **hairpin,** which slows the elongation complex. The elongation complex then dissociates as it reaches the A:T-rich area.

In eukaryotes, mRNA synthesis, catalyzed by pol II, proceeds along the DNA template until a **polyadenylation signal** (polyA site) is encountered. At this point, the process of termination of transcription is activated. There is no specific sequence in DNA that specifies termination of transcription. As the polymerase proceeds past the polyA site, the nascent mRNA is released by an endonuclease associated with the carboxy terminal end of pol II. RNA synthesized beyond the site trails out of the polymerase and is bound by another exonuclease that begins to degrade the RNA 5′ to 3′ toward the RNA polymerase. When the exonuclease catches up with the polymerase, transcription stops.

Pol I in eukaryotes terminates transcription just prior to a site in the DNA (Sal box) with the cooperation of a termination factor, TTF1.[51] The pol III termination signal is a run of adenine residues in the template. Pol III transcription termination requires a termination factor.

TYPES/STRUCTURES OF RNA

There are several types of RNAs found in the cell. Ribosomal RNA, mRNA, transfer RNA, and small nuclear RNAs have distinct cellular functions.

Ribosomal RNA

The largest component of cellular RNA is **ribosomal RNA (rRNA),** comprising 80% to 90% of the total cellular RNA. The various types of ribosomal RNAs are named for their sedimentation coefficient (S) in density-gradient centrifugation.[52] rRNA is an important structural and functional part of the ribosomes, cellular organelles where proteins are synthesized (Fig. 1.27).

In prokaryotes, there are three rRNA species, the 16S found in the ribosome small subunit and the 23S and 5S found in the ribosome large subunit, all synthesized from the same gene. In eukaryotes, rRNA is synthesized from highly repeated gene clusters. Eukaryotic rRNA is copied from DNA as a single 45S precursor RNA (preribosomal RNA) that is subsequently processed into the 18S species of the ribosome small subunit and 5.8S and 28S species of the large subunit. Another rRNA species, 5S, found in the large ribosome subunit in eukaryotes, is synthesized separately.

Messenger RNA

Messenger RNA (mRNA) is the initial connection between the information stored in DNA and the translation apparatus that will ultimately produce the protein products responsible for the phenotype. In prokaryotes, mRNA is synthesized and simultaneously translated into protein. Prokaryotic mRNA is sometimes polycistronic; that is, one mRNA codes for more than one protein. Eukaryotic mRNA, in contrast, is monocistronic, having only one protein per mRNA. Eukaryotes can, however, produce different proteins from the same DNA sequences by starting the RNA synthesis in different places or by processing the mRNA differently. In eukaryotes, copying of RNA from DNA and protein synthesis from the RNA are separated by the nuclear membrane barrier. Eukaryotic mRNA undergoes a series of post-transcriptional processing events before it is translated into protein (Fig. 1.28).

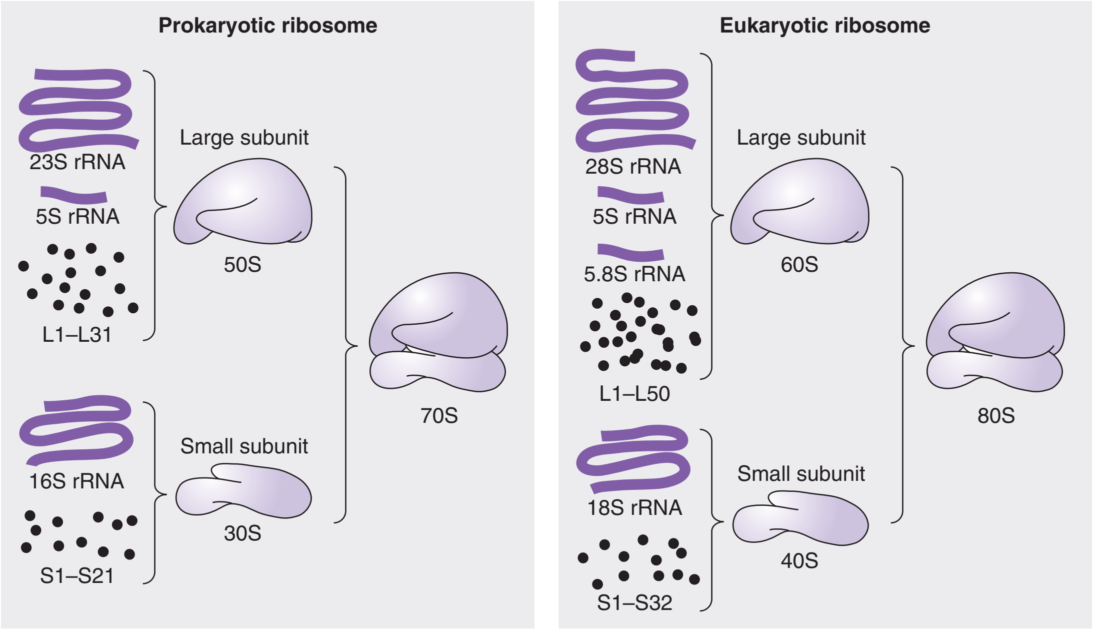

Prokaryotic ribosome
23S rRNA
5S rRNA
L1–L31
Large subunit
50S
70S
16S rRNA
S1–S21
Small subunit
30S
Eukaryotic ribosome
28S rRNA
5S rRNA
5.8S rRNA
L1–L50
Large subunit
60S
80S
18S rRNA
S1–S32
Small subunit
40S

FIGURE 1.27 Prokaryote and eukaryote ribosomal subunits are of similar structure but different size. Ribosomal RNAs *(left)* are assembled with 52 or 82 ribosomal proteins *(center)* to make the subunits that will form the complete ribosome in association with mRNA.

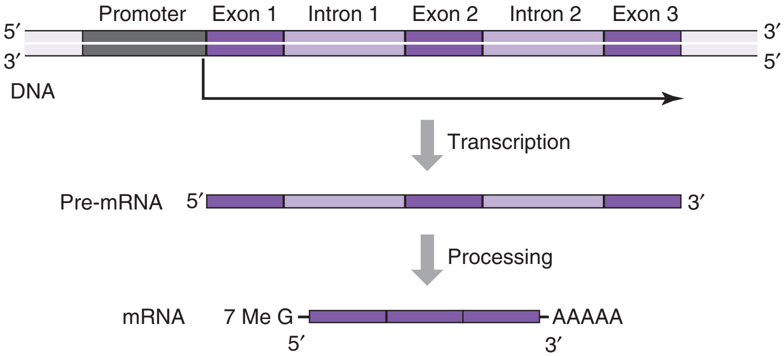

Promoter
Exon 1
Intron 1
Exon 2
Intron 2
Exon 3
5′
3′
3′
5′
DNA
Transcription
Pre-mRNA
5′
3′
Processing
mRNA
7 Me G
AAAAA
5′
3′

FIGURE 1.28 DNA *(top)* and heteronuclear RNA *(middle)* contain intervening (intron) and expressed (exon) sequences. The introns are removed, and the mature RNA is capped and polyadenylated *(bottom)*.

The amount of a particular mRNA in a cell is related to the requirement for its final product. Some messages are transcribed constantly and are relatively abundant in the cell (**constitutive** transcription), whereas others are transcribed only at certain times during the cell cycle or under particular conditions (**inducible,** or regulatory, transcription). Even the most abundantly transcribed mRNAs are much less plentiful in the cell than rRNA.

Advanced Concepts

The secondary structure of rRNA is important for the integrity and function of the ribosome. Not only is it important for ribosomal structure, but it is also involved in the correct positioning of the ribosome on the mRNA and with the transfer RNA during protein synthesis.[53,54]

Messenger RNA Processing

Polyadenylation

The study of mRNA in eukaryotes was facilitated by the discovery that most messengers carry a sequence of **polyadenylic acid** at the 3′ terminus, a **polyA tail.** The run of adenines was first discovered by hydrogen bonding of mRNA to polydeoxythymine on poly(dT) cellulose.[55] Polyuridine or **polythymine** residues covalently attached to cellulose or sepharose substrates are often used to specifically isolate mRNA in the laboratory.

The polyA tail is not coded in genomic DNA. It is added to the RNA after synthesis of the pre-mRNA. A protein complex recognizes the RNA sequence, AAUAAA, and cleaves the RNA chain 11 to 30 bases 3′ to that site. The enzyme that cuts pre-mRNA in advance of **polyadenylation** has not been identified. Recent studies suggest that a component of the protein complex related to the system that is responsible for removing the 3′ extension from pretransfer RNAs may also be involved in the generation of the 3′ ends on mRNA.[56,57] The enzyme **polyadenylate polymerase** is responsible for adding the adenines to the end of the transcript. A run of up to 200 nucleotides of polyA is typically found on mRNA in mammalian cells.

Historical Highlights

A DNA copy (**complementary DNA,** copy DNA [cDNA]) of mature mRNA can be made by reverse transcription of mRNA (synthesis of DNA from the mRNA template). Compared with the original gene on the chromosome, the cDNA version of eukaryotic genes is smaller than the chromosomal version. Restriction enzyme mapping confirms that this is not due to premature termination of the genomic transcript. The entire functional gene is present in the shorter sequence because cDNA versions of eukaryotic genes can be expressed (transcribed and translated) into complete proteins. The chromosomal version of the gene must, therefore, have extra sequences inserted between the protein-coding sequences. The direct location and size of these intervening sequences or introns were first demonstrated by electron microscopy of hybrids between mRNA and cDNA[58] using the method of R-loop mapping developed by White and Hogness.[59] In these experiments, mRNA and duplex genomic DNA were incubated together at conditions under which the more stable RNA–DNA hybrids are favored over the DNA duplexes. The resulting structures released loops of unpaired DNA (introns). The DNA duplexed with RNA corresponded to the coding sequences (exons).

Capping

Eukaryotic mRNA is blocked at the 5′ terminus by a 5′-5′ pyrophosphate bridge to a methylated guanosine.[60] The structure is called a **cap.** The cap is a 5′-5′ pyrophosphate linkage of 7-methyl guanosine to either 2′ *O*-methyl guanine or 2′ *O*-methyl adenine of the mRNA:

$$^{7\text{-methyl}}\mathrm{G}^{5'} + \mathrm{pNpNp}^{5'} + {}^{2'O\text{-methyl}}(\mathrm{G\ or\ A})\mathrm{pNpNpNp}\ldots$$

where p represents a phosphate group, and N represents any nucleotide.

Advanced Concepts

About 30% of the mRNAs, notably histone mRNAs, are not polyadenylated.[61] The functional differences between polyA+ and polyA− mRNA are not clear. The polyA tail may be involved in the movement of the mRNA from the nucleus to the cytoplasm, association with other cell components, the maintenance of secondary structure, or the stability of the molecule. Abnormalities in the 3′-end processing mechanism have been found in oncological, immunological, neurological, and hematological diseases.[62]

The cap confers a protective function and serves as a recognition signal for the translational apparatus. Caps differ with respect to the methylation of the end nucleotide of the mRNA. In some cases, 2′ *O*-methylation occurs not only on the first but also on the second nucleotide from the cap. Other caps methylate the first three nucleotides of the RNA molecule.

Splicing

Prokaryotic structural genes contain uninterrupted lengths of open reading frame, sequences that code for amino acids. In contrast, eukaryotic coding regions are interrupted with long stretches of noncoding DNA

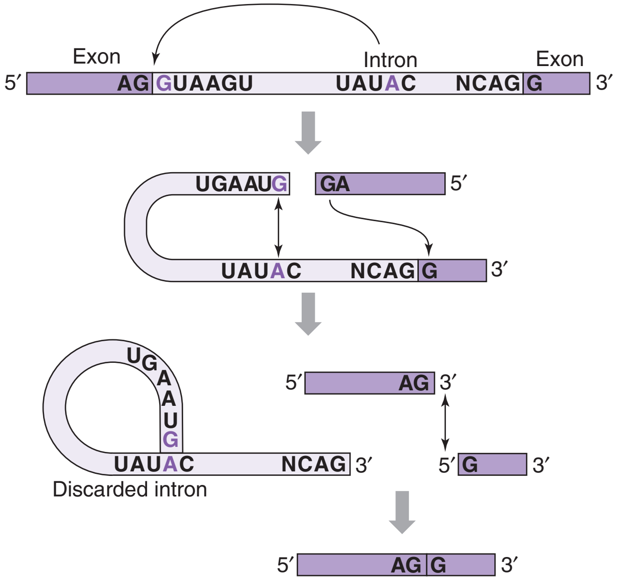


FIGURE 1.29 RNA splicing at the 5′ splice site (AGGUAAGU), branch (UAUAC), and 3′ splice site (NCAGG) consensus sequences. The intron is removed through a transesterification reaction involving a guanine nucleotide of the 5′ site and an adenine in the branch sequence. The product of this reaction is the discarded intron in a lariat structure. Another transesterification reaction connects the exons.

(intervening) sequences called **introns.** Newly transcribed mRNA, heteronuclear RNA (hnRNA), is much larger than mature mRNA because it still contains the intervening sequences. Labeling studies demonstrated that the hnRNA is capped and tailed and that these modifications survive the transition from hnRNA to mRNA, which is simply a process of removing the intervening sequences from the hnRNA. Introns are removed from hnRNA by **splicing** (Fig. 1.29). The remaining sequences that code for the protein product are **exons.**

Historical Highlights

Although removal of all nuclear introns requires protein catalysts, some introns are removed without the participation of protein factors in a self-splicing reaction. The discovery of self-splicing was the first demonstration that RNA could act as an enzyme.[63,64]

Inspection of splice junctions from several organisms and genes has demonstrated the following consensus sequences for the donor and acceptor splice junctions of group I, group II, and nuclear introns:[65]

(intron)
branch site donor site acceptor site
A G // ***G U*** *A A G U Y N C U R* ***A*** *C* Y_N *N C* ***A G*** // G

The branch-point sequence YNCURAC is variable in mammals but almost invariant in the yeast *Saccharomyces cerevisiae* (UACUAAC).

Advanced Concepts

Caps are present on all eukaryotic mRNA bound for translation, with the exception of some mRNA transcribed from mitochondrial DNA. Capping occurs after initiation of transcription, catalyzed by the enzyme guanylyltransferase. This enzyme links a guanosine monophosphate provided by guanosine triphosphate to the 5′ phosphate terminus of the RNA with the release of **pyrophosphate.** In some viruses, guanosine diphosphate provides the guanidine residue, and monophosphate is released. Caps of mRNA are recognized by ribosomes just before translation.[66]

There are four types of introns—group I, group II, nuclear, and tRNA—depending on the mechanism of their removal from hnRNA. Group I introns are found in nuclear, mitochondrial, and chloroplast genes. Group II introns are found in mitochondrial and chloroplast genes. Group I introns require a guanosine-triphosphate molecule to make a nucleophilic attack on the 5′ phosphate of the 5′ end of the intron. This leaves a 3′ OH at the end of the 5′ exon (splice donor site), which attacks the 5′ end of the next exon (splice acceptor site), forming a new phosphodiester bond and releasing the intervening sequence. Group II introns are removed in a similar reaction initiated by the 2′ OH of an adenosine within the intron attacking the 5′ phosphate at the splice donor site. When the 3′ OH of the splice donor site bonds with the splice acceptor site of the next exon, the intervening sequence is released as a lariat structure (see Fig. 1.29).

This lariat contains an unusual 2′, 3′, and 5′ triply linked nucleotide, the presence of which proved the mechanism. Removal of nuclear introns occurs by the same transesterification mechanism, except this reaction is catalyzed by specialized RNA–protein complexes (small nucleoprotein particles). These complexes contain the small nuclear RNAs U1, U2, U4, U5, and U6.

Why are eukaryotic genes interrupted by introns? Splicing may be important for the timing of the translation of mRNA in the cytoplasm. One theory holds that introns evolved as a means of increasing recombination frequency within genes as well as between genes without breaking coding sequences. The discontinuous nature of eukaryotic genes may also protect the coding regions from genetic damage by toxins or radiation.

Alternative splicing modifies products of genes by alternate insertion of different exons. For example, the production of calcitonin in the thyroid or calcitonin gene-related peptide in the brain depends on the exons included in the mature mRNA in these tissues.[67] Alternative splicing has been found in over 40 different genes.

Advanced Concepts

The splicing of transfer RNA (tRNA) transcripts involves breakage and reunion of the RNA chain. Endonucleases cleave the tRNA precisely at the intron ends. The resulting tRNA ends, a 2′, 3′ cyclic phosphate and a 5′ OH, are then ligated in a complex reaction that requires ATP, followed by further base modification in some tRNAs.

Abnormalities in the splicing process are responsible for several disease states. Some β-thalassemias result from mutations in the splice recognition sequences of the β-globin genes. Certain autoimmune conditions result from production of antibodies to RNA–protein complexes. Autoantibodies against U1 RNA, one of the small nuclear RNAs required for splicing, are associated with systemic lupus erythematosus.

Small Nuclear RNA

Small nuclear RNA (snRNA) functions in splicing in eukaryotes. Small nuclear RNA stays in the nucleus after its transcription by RNA polymerase I or III. These RNAs sediment in a range of 6 to 8S. Small nuclear RNAs isolated from hepatoma and cervical carcinoma cell lines are summarized in Table 1.5.

TABLE 1.5 Small Nuclear RNA Isolated From HeLa Cervical Carcinoma and Novikoff Hepatoma Cells[17]

Species (HeLa)	Species (Novikoff)	Approximate Length (Bases)
SnA	U5	180
SnB		210
SnC	U2	196
SnD	U1B	171
SnE/ScE (5.8S rRNA)	5.8S	
SnF	U1A	125
SnG/ScG (5S rRNA)	5S I and II	
SnG′	5S III	120
SnH	4.5S I, II, and III	96
SnI (tRNA)		
SnK		260
SnP		130
ScL (viral 7S)		260
ScM		180
ScD		180

Transfer RNA

Translation of information from nucleic acid to protein requires reading of the mRNA by **ribosomes,** using adaptor molecules or **transfer RNA** (tRNA). Transfer RNAs are relatively short, single-stranded polynucleotides of 73 to 93 bases in length, MW 24,000 to 31,000. There is at least one tRNA for each amino acid.

Eight or more of the nucleotide bases in all tRNAs are modified, usually methylated, after the tRNA synthesis.

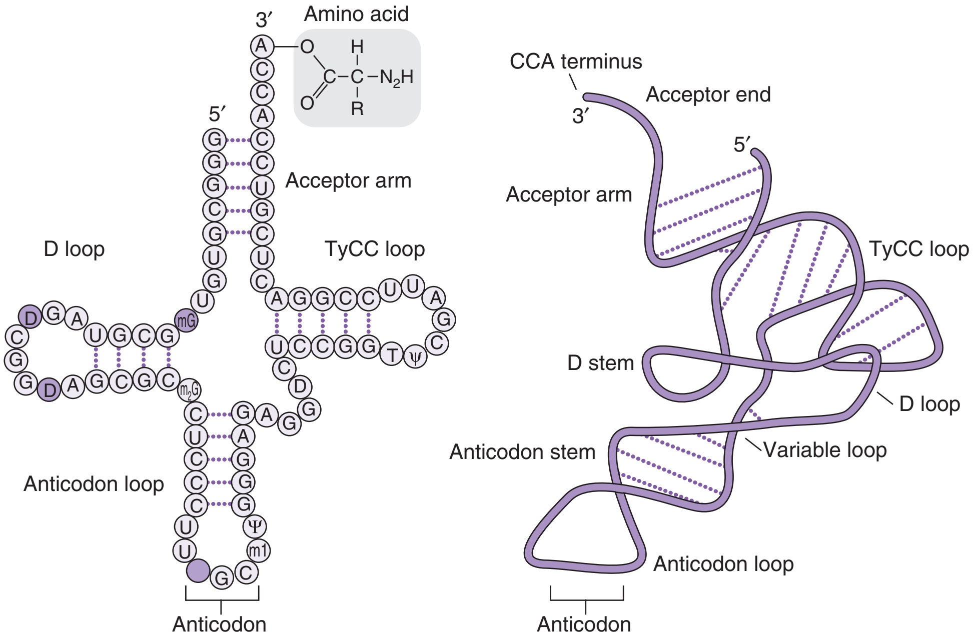


FIGURE 1.30 Alanine tRNA is an example of the general structure of tRNA, which is often depicted in a cruciform structure *(left)*. The inverted "L" *(right)* more accurately depicts the structure formed by intrastrand hydrogen bonding.

Most tRNAs have a guanylic residue at the 5′ end and the sequence CCA at the 3′ end. Through intrastrand hybridization, tRNAs take on a cruciform structure of four to five double-stranded stems and three to four single-stranded loops (Fig. 1.30). The CCA at the 3′ end of the tRNA is where the amino acid will be covalently attached to the tRNA. A seven-base loop (the TΨC loop, where Ψ stands for the modified nucleotide pseudouridine) contains the sequence 5′-TΨCG-3′. The variable loop is larger in longer tRNAs. Another seven-base loop (the anticodon loop) contains the three-base-pair anticodon that is complementary to the mRNA codon of its cognate amino acid. An 8- to 12-base loop (D loop) is relatively rich in dihydrouridine, another modified nucleotide. Studies of pure crystalline tRNA using x-ray diffraction reveal that the cruciform secondary structure of tRNA takes on an additional level of hydrogen bonding between the D loop and the TΨC loop to form a γ-shaped structure (see Fig. 1.30).

Advanced Concepts

Small nuclear RNAs serve mostly a structural role in the processing of mRNA. Several of a family of proteins (Sm proteins) assemble into a (60 Å by 30 to 40 Å) doughnut-shaped complex that interacts with the U-rich regions of poly (U) RNAs.[68,69]

U1 RNA is complementary to sequences at the splice donor site, and its binding distinguishes the sequence GU in the splice site from other GU sequences in the RNA. U2 RNA recognizes the splice acceptor site. In lower eukaryotes, another protein binds to the branch-point sequence, initiating further protein assembly and association of U4, U5, and U6, with the looped RNA forming a complex, the spliceosome, in which the transesterification reaction linking the exons together takes place.

Advanced Concepts

Mitochondria contain distinct, somewhat smaller, tRNAs.

Advanced Concepts

The tRNA genes contain 14 to 20 extra nucleotides in the sequences coding for the anticodon loop that are transcribed into the tRNA. Enzymes that recognize other tRNA modifications remove these sequences (introns) by a cleavage-ligation process. Intron removal, the addition of CCA to the 3′ end, and nucleotide modifications all occur following tRNA transcription. Enzymatic activities responsible for intron removal and the addition of CCA may also contribute to intron removal and polyA addition to mRNA.

Historical Highlights

In 1964, Robert Holley and colleagues at Cornell University solved the first tRNA sequence. The sequence was that of alanine tRNA of yeast.[70] Yeast tRNA^{ala} is 76 bases long; 10 of these bases are modified.

Other RNAs

Since the late 1990s, increasing varieties of RNA species have been described in prokaryotes and eukaryotes. In addition to RNA synthesis and processing, these molecules influence numerous cellular processes, including plasmid replication, bacteriophage development, chromosome structure, and development. Untranslated RNA molecules have been termed sRNAs in bacteria and noncoding RNAs (ncRNAs) in eukaryotes. Their role in controlling phenotype will be discussed in Chapter 2.

RNA POLYMERASES

RNA synthesis is catalyzed by RNA polymerase enzymes (Table 1.4). One multisubunit prokaryotic enzyme is responsible for the synthesis of all types of RNA in the prokaryotic cell. Eukaryotes have three different RNA polymerase enzymes. All of these enzymes are DNA-dependent RNA polymerases; that is, they require a DNA template. In contrast, RNA-dependent **RNA polymerases** require an RNA template.

Bacterial RNA polymerase consists of five subunits, two α and one of each β, β', and σ (Fig. 1.31).[71]

Historical Highlights

The $\alpha 2$, β, β' core enzyme retains the catalytic activity of the $\alpha 2$, β', σ complete enzyme, or holoenzyme, suggesting that the sigma factor plays no role in RNA elongation.[72] In fact, the sigma factor is released at RNA initiation. The role of the sigma factor is to guide the complete enzyme to the proper site of initiation on the DNA.

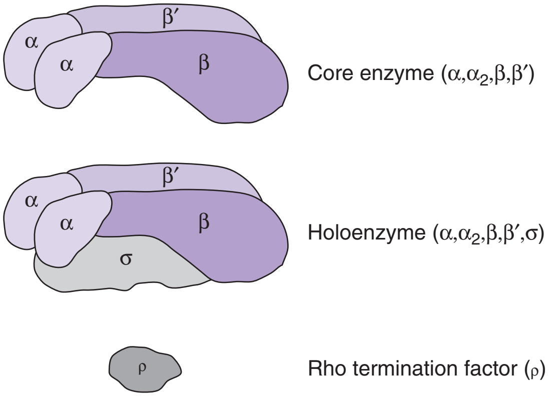


FIGURE 1.31 Prokaryotic RNA polymerase is made up of separate proteins. The four subunits that make up the core enzyme have the capacity to synthesize RNA. The sigma cofactor aids in the accurate initiation of RNA synthesis. The rho cofactor aids in termination of RNA synthesis.

Advanced Concepts

The nucleotide sequence of mRNA can be altered after RNA synthesis by RNA editing. There are two mechanisms of RNA editing in humans. In one, such enzymes as cytidine deaminase and adenosine deaminase convert C to U and A to inosine (read as G), respectively. These enzymes recognize specific sequences in the mRNA. The other mechanism involves guide RNA (gRNA), a small nuclear RNA that hydrogen bonds to the mRNA transcript and mediates the addition or removal of bases from the mRNA. The former substitution mechanism is responsible for RNA editing of the *APOB* gene so that different versions are made in the liver and intestines.

In eukaryotes, there are three multisubunit RNA polymerases, RNA polymerase I, II, and III (Table 1.5), and a single-subunit mitochondrial RNA polymerase imported to organelles. The three RNA polymerases in eukaryotic cells were first distinguished by their locations in the cell. RNA polymerase I (pol I) is found in the nucleolus. RNA polymerase II (pol II) is found in the nucleus. RNA polymerase III (pol III), one of the first nucleic acid polymerases discovered, is also found in the nucleus and sometimes in the cytoplasm.[73]

Advanced Concepts

The three polymerases were also distinguished by their differential sensitivity to the toxin α-amanitin.[74,75] This toxin is isolated from the poisonous mushroom *Amanita phalloides*. Pol II is sensitive to relatively low amounts of this toxin, pol III is sensitive to intermediate levels, and pol I is resistant. α-Amanitin was invaluable in the research setting to determine which polymerase activity is responsible for the synthesis of newly discovered types of RNA and to dissect the biochemical properties of the polymerases.

Advanced Concepts

The most well-studied eukaryotic RNA polymerase is pol II from the yeast *S. cerevisiae*. It is a 0.4-megadalton complex of 12 subunits. The yeast enzyme works in conjunction with a large complex of proteins required for promoter recognition and melting, transcription initiation, elongation and termination, and transcript processing (splicing, capping, and polyadenylation).

RNA viruses carry their own RNA-dependent RNA polymerases. Hepatitis C virus, Zika virus, and Dengue virus carry this type of polymerase to replicate their RNA genomes. RNA-dependent RNA polymerase activity has also been found in plants and lower eukaryotes, where the purpose of these enzymes in cells may be associated with gene silencing.[76] The enzyme activity may serve as a therapeutic target for antiviral agents.[77]

PolyA polymerase is a template-independent RNA polymerase. This enzyme adds adenine nucleotides to the 3′ end of mRNA.[78] The resulting polyA tail is important for mRNA stability and translation into protein.

OTHER RNA-METABOLIZING ENZYMES

Ribonucleases

Ribonucleases degrade RNA in a manner similar to the degradation of DNA by deoxyribonucleases (Table 1.6). There are several types of ribonucleases, generally classified as endoribonucleases and exoribonucleases.

Endonucleases include RNase H, which digests the RNA strand in a DNA–RNA hybrid; RNase I, which cleaves single-stranded RNA; and RNase III, which digests double-stranded RNA. RNase P removes precursor nucleotides from tRNA molecules. RNase A, RNase T1, and RNase T2 cleave single-stranded RNA at specific residues. A combination of RNase A, T1, and T2 is used in some laboratory procedures investigating gene expression and transcript structure. An endoribonuclease formed by cleavage and polyadenylation-specific factor (CPSF) and other factors that bind to the polyA site are required for proper termination of RNA synthesis in

TABLE 1.6 RNases Used in Laboratory Procedures[82]

Enzyme	Source	Type	Substrate
RNase A	Bovine	Endoribonuclease	Single-stranded RNA 3′+ to pyrimidine residues
RNase T1	*Aspergillus*	Endoribonuclease	3′ Phosphate groups of guanines
RNase H	*E. coli*	Exoribonuclease	RNA hybridized to DNA
RNase CL3	*Gallus*	Endoribonuclease	RNA next to cytidylic acid
Cereus RNase	*Physarum*	Endoribonuclease	Cytosine and uracil residues in RNA
RNase Phy M	*Physarum*	Endoribonuclease	Uracil, adenine, and guanine residues in RNA
RNase U2	*Ustilago*	Endoribonuclease	3′ Phosphodiester bonds next to purines
RNase T2	*Aspergillus*	Endoribonuclease	All phosphodiester bonds, preferably next to adenines
S1 nuclease	*Aspergillus*	Exoribonuclease	RNA or single-stranded DNA
Mung bean nuclease	Mung bean sprouts	Exoribonuclease	RNA or single-stranded DNA
RNase Phy I	*Physarum*	Exoribonuclease	Guanine, adenine, or uracil residues in RNA

mammals.[79] Along with RNA polymerase II subunits and other proteins, this activity cuts the nascent RNA transcript before addition of the polyA tail by polyA polymerase. The CPSF complex may also play a role in alternative splicing.[80] Exoribonucleases digest single-stranded RNA from the 3′ or 5′ ends. This group of enzymes includes polynucleotide phosphorylase (PNPase), RNase PH, RNase II, RNase R, RNase D, RNase T, and others. Like the endoribonuclease RNase P, RNase D is involved in the 3′ to 5′ processing of pre-tRNAs.

RNases are ubiquitous, stable enzymes that degrade all types of RNA. Some RNases are secreted by higher eukaryotes, possibly as an antimicrobial defense mechanism.[81,82] They are of particular concern in laboratories where RNA work is performed because they are very resistant to inactivation. Special precautions are used to protect RNA from degradation by these enzymes (see Chapter 15, "Quality Assurance and Quality Control in the Molecular Laboratory").

RNA Helicases

RNA synthesis and processing require the activity of helicases to catalyze the unwinding of double-stranded RNA. These enzymes have been characterized in prokaryotic and eukaryotic organisms. Some RNA helicases work exclusively on RNA. Others can work on DNA:RNA heteroduplexes and DNA substrates. Another activity of these enzymes is in the removal of proteins from RNA–protein complexes.

PROTEINS AND THE GENETIC CODE

Proteins are the products of transcription and translation of the nucleic acids. Even though nucleic acids are most often the focus of "molecular analysis," the ultimate effect of the information stored and delivered by the nucleic acid is manifested in proteins. In the medical laboratory, analysis of the amount and mutational status of specific proteins has long been performed in situ using immunohistochemistry, on live cells using flow cytometry, and on isolated proteins by enzyme-linked immunosorbent assays, **capillary electrophoresis,** and western blots. More recently, global protein analysis by mass spectrometry (proteomics) has also been applied to clinical work. Even if proteins are not being tested directly, they manifest the phenotype directed by the nucleic acid information. In order to interpret the results of nucleic acid analysis accurately, therefore, it is important to understand the movement of genetic information from DNA to protein as dictated by the **genetic code.**

Amino Acids

Proteins are polymers of **amino acids.** Each amino acid has characteristic biochemical properties determined by the nature of its amino acid **side chain** (Fig. 1.32). Amino acids are grouped according to their polarity (tendency to interact with water at pH 7) as follows: **nonpolar,** uncharged **polar,** negatively charged polar, and positively charged polar (Table 1.7).

The properties of amino acids that make up a protein determine the shape and biochemical nature of the protein. A single protein can have separate **domains** with different properties. For example, transmembrane proteins have several stretches of hydrophobic amino acids positioned in the lipid membrane of the cell, and they might also have hydrophilic or charged **extracellular** domains (Fig. 1.33).

Amino acids are synthesized in vivo by stereospecific enzymes so that naturally occurring proteins are made of amino acids of L-stereochemistry. The central asymmetric carbon atom of the amino acid is attached to a carboxyl group, an amino group, a hydrogen atom, and the side chain. Proline differs from the rest of the amino acids in that its side chain is cyclic, with the amino group attached to the end carbon of the side chain making a five-carbon ring (see Fig. 1.32).

Amino acids are also classified by their biosynthetic origins or similar structures based on a common biosynthetic precursor (Table 1.8). Histidine has a unique synthetic pathway using metabolites common to purine nucleotide biosynthesis, which affords the connection of amino acid synthesis to nucleotide synthesis.

At pH 7, most of the carboxyl groups of the amino acids are ionized, and the amino groups are not. The ionization can switch between the amino and carboxyl groups, making the amino acids **zwitterions** at physiological pH (Fig. 1.34). At certain pH levels, amino acids will become completely positively or negatively charged. These are the pK values for each amino acid. At the pH where an amino acid is neutral, its positive and negative charges are in balance. This is the pI value. Each amino acid will have its characteristic pI. The pI of a **peptide** or protein is determined by the ionization state (positive and negative charges) of the side chains of its constituent amino acids.

The amino and carboxyl terminal groups of the amino acids are joined in a carbon-carbon-nitrogen (–C–C–N–) substituted amide linkage **(peptide bond)** to form the protein backbone (Fig. 1.35). Two amino acids joined together by a peptide bond make a dipeptide. Peptides with additional units are tri-, tetra-, pentapeptides, and so forth, depending on how many units are attached to each other. At one end of the peptide will be an amino group (the **amino-terminal,** or NH_2 end), and at the opposite terminus of the peptide will be a carboxyl group (the **carboxy-terminal,** or COOH end). Like the 5′ to 3′ direction of nucleic acids, peptide chains grow from the amino to the carboxy terminus.

Proteins are **polypeptides** that can reach sizes of more than a thousand amino acids in length. The information stored in the sequence of nucleotides in DNA is transcribed and translated into an amino acid sequence that will ultimately bring about the genetically coded phenotype.

Proteins constitute the most abundant macromolecules in cells. The collection of proteins encoded in all of an organism's DNA is a **proteome.** The proteome of humans is larger than the genome (collection of all genes), possibly 10 times its size, although the exact number of proteins is difficult to assess.[83] This is because a single gene can give rise to more than one protein through alternate splicing and other post-transcriptional/post-translational modifications.

Advanced Concepts

There are two non-canonical amino acids, pyrrolysine and **selenocysteine.** Pyrrolysine is found in Archebacteria. Selenocysteine is a component of **selenoproteins,** such as glutathione peroxidase and formate dehydrogenase. Pyrrolysine is encoded by UAG and selenocysteine by UGA where they are inserted instead of a termination signal. In humans, changes in the translation of UGA can lead to symptoms of selenium deficiency.[84]

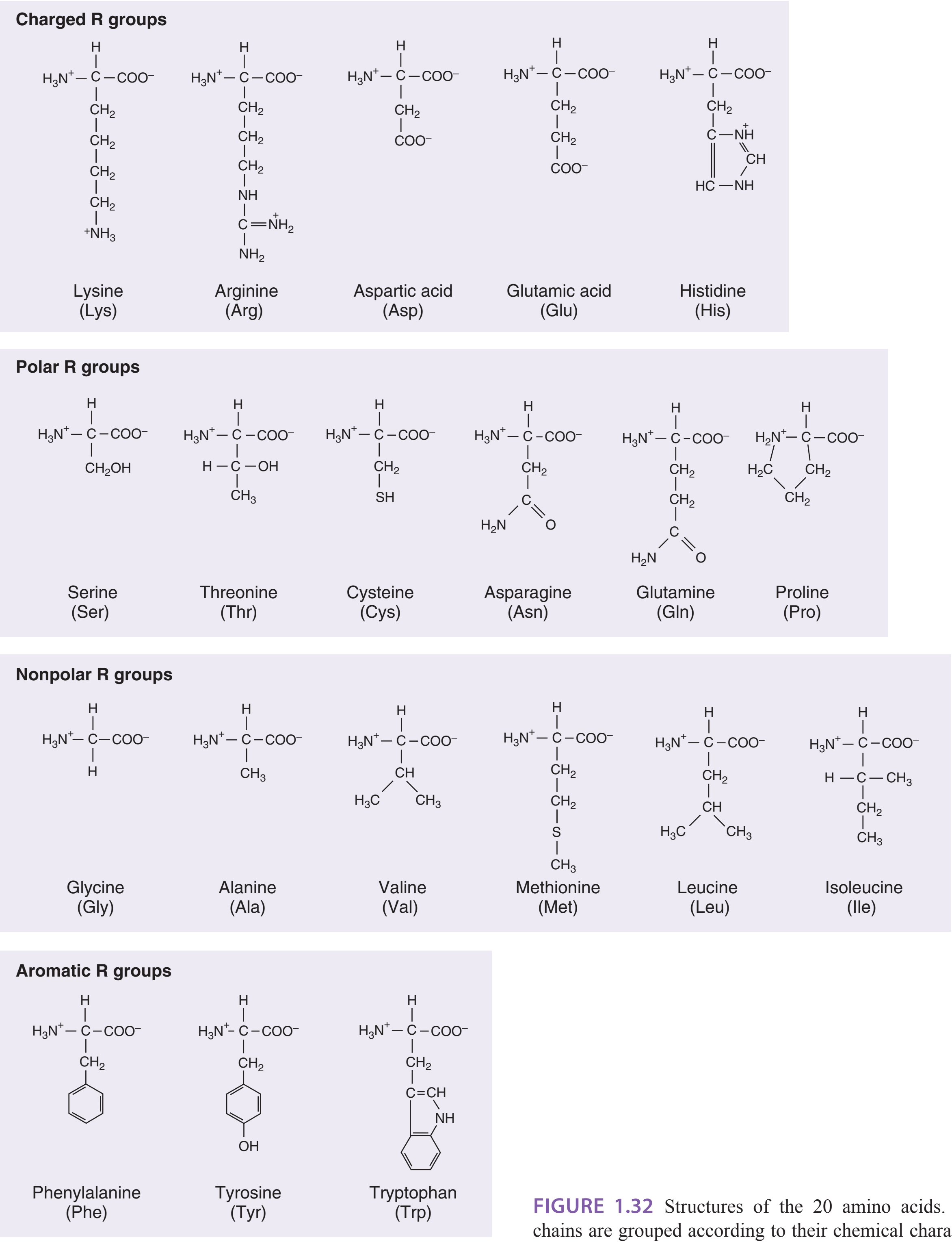

Charged R groups
Lysine
(Lys)
Arginine
(Arg)
Aspartic acid
(Asp)
Glutamic acid
(Glu)
Histidine
(His)
Polar R groups
Serine
(Ser)
Threonine
(Thr)
Cysteine
(Cys)
Asparagine
(Asn)
Glutamine
(Gln)
Proline
(Pro)
Nonpolar R groups
Glycine
(Gly)
Alanine
(Ala)
Valine
(Val)
Methionine
(Met)
Leucine
(Leu)
Isoleucine
(Ile)
Aromatic R groups
Phenylalanine
(Phe)
Tyrosine
(Tyr)
Tryptophan
(Trp)

FIGURE 1.32 Structures of the 20 amino acids. The side chains are grouped according to their chemical characteristics.

TABLE 1.7 Classification of Amino Acids Based on Polarity of Their Side Chains

Classification	Amino Acid	Abbreviations
Nonpolar	Alanine	Ala, A
	Isoleucine	Ile, I
	Leucine	Leu, L
	Methionine	Met, M
	Phenylalanine	Phe, F
	Tryptophan	Trp, W
	Valine	Val, V
Polar	Asparagine	Asn, N
	Cysteine	Cys, C
	Glutamine	Gln, Q
	Glycine	Gly, G
	Proline	Pro, P
	Serine	Ser, S
	Threonine	Thr, T
	Tyrosine	Tyr, Y
Negatively charged (acidic)	Aspartic acid	Asp, D
	Glutamic acid	Glu, E
Positively charged (basic)	Arginine	Arg, R
	Histidine	His, H
	Lysine	Lys, K

TABLE 1.8 Amino Acid Biosynthetic Groups

Biosynthetic Group	Precursor	Amino Acids
α-Ketoglutarate group	α-Ketoglutarate	Gln, Q Glu, E Pro, P Arg, R
Pyruvate group	Pyruvate	Ala, A Val, V Leu, L
Oxalate group	Oxaloacetic acid	Asp, D Asn, N Lys, Q Ile, I Thr, T Met, M
Serine group	3-Phosphoglycerate	Gly, G Ser, S Cys, C
Aromatic group (Unique biosynthesis)	Chorismate	Phe, F Trp, W Tyr, Y His, H

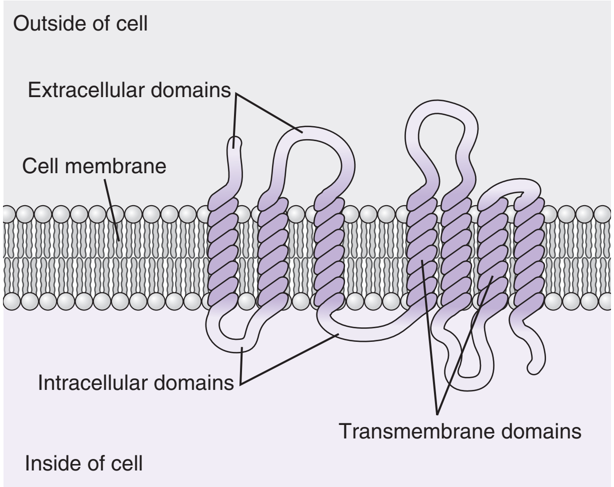


FIGURE 1.33 Transmembrane proteins have hydrophobic transmembrane domains and hydrophilic domains exposed to the intracellular and extracellular spaces. The biochemical nature of these domains results from their distinct amino acid compositions.

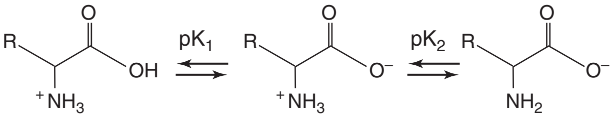


FIGURE 1.34 An amino acid is positively charged at pK_1 and negatively charged at pK_2. At the pH where the positive and negative charges balance (pI), the molecule is neutral.

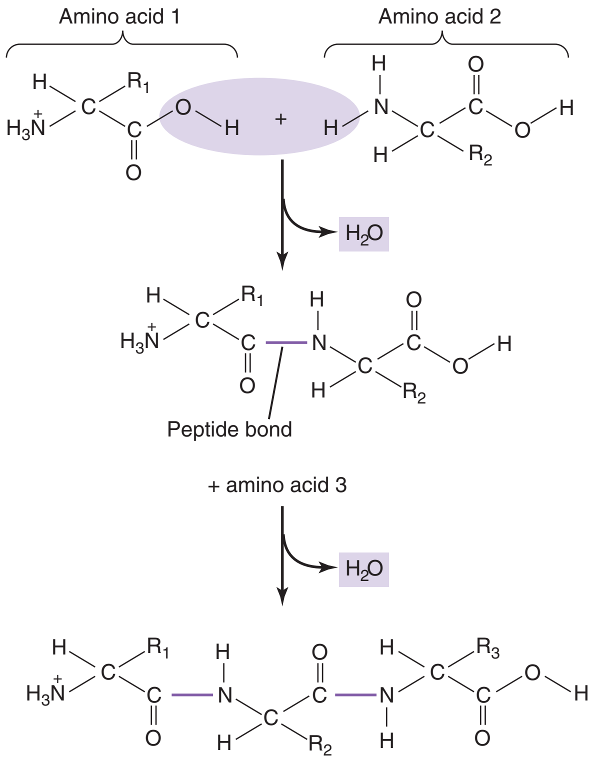


FIGURE 1.35 The peptide bond is a covalent linkage of the carboxyl C of one amino acid with the amino N of the next amino acid. One molecule of water is released in the reaction.

Historical Highlights

The primary sequence of proteins can be determined by a method first described in Fred Sanger's report of the amino acid sequence of insulin.[85,86] This procedure was carried out in six steps. First, the protein was dissociated into amino acids. The dissociation products were separated by ion-exchange chromatography in order to determine the type and amount of each amino acid. Second, the amino terminal and carboxy terminal amino acids were determined by labeling with 1-fluoro-2,4-dinitrobenzene and digestion with carboxypeptidase, respectively. The complete protein was then fragmented selectively at lysines and arginines with trypsin to 10 to 15 amino acid peptides. Fourth, Edmund degradation with phenylisothiocyanate and dilute acid labeled and removed the amino terminal residue. This was repeated on the same peptide until all the amino acids were identified. Fifth, after another selected cleavage of the original protein using cyanogen bromide, chymotrypsin, or pepsin and identification of the peptides by chromatography, the peptides were again sequenced using the Edmund degradation. Sixth, the complete amino acid sequence could be assembled by identification of overlapping regions.

The sequence of amino acids in a protein determines the nature and activity of that protein. This sequence is the **primary structure** of the protein and is read by convention from the amino terminal end to the carboxy terminal end. Minor changes in primary structure can alter the activity of proteins dramatically. The single amino acid substitution that produces hemoglobin S in sickle cell anemia is a well-known example. Replacement of a soluble glutamine residue with a hydrophobic valine at the sixth residue changes the nature of the protein so that it packs aberrantly in corpuscles and drastically alters cell shape. Minor changes in primary structure can have such drastic effects because the amino acids must often cooperate with one another to bring about protein structure and function.

Advanced Concepts

Even small peptides can have biological activity. Hormones such as insulin, glucagons, corticotropin, oxytocin, bradykinin, and thyrotropin are examples of peptides (9 to 40 amino acids long) with strong biological activity. Several antibiotics, such as penicillin and streptomycin, are also peptides.

Advanced Concepts

Protein sequence can be inferred from DNA sequence, although the degeneracy of the genetic code will result in several possible protein sequences for a given DNA sequence. Many databases on the frequency of codon usage in various organisms are available.[87,88] These can be used to more accurately predict amino acid sequences from DNA sequence data.

Interactions between amino acid side chains fold a protein into predictable configurations. These include ordered beta or beta-pleated sheets and less-ordered alpha helices, or random coils. The alpha-helix and beta-sheet structures in proteins (Fig. 1.36) were first described by Linus Pauling and Robert Corey in 1951.[89-92] This level of organization is the **secondary structure** of the protein. Some proteins, especially structural proteins, consist almost entirely of alpha helices or beta sheets. Globular proteins have varying amounts of alpha helices and beta sheets.

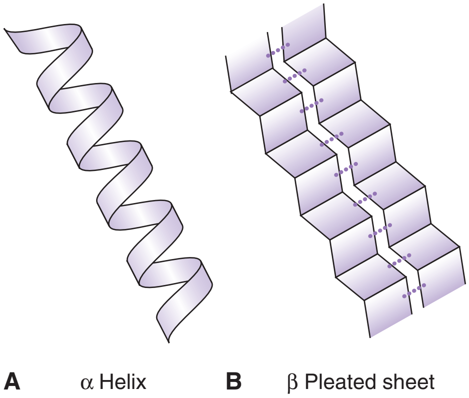


FIGURE 1.36 Secondary structure of proteins includes the alpha helix (A) and the beta pleated sheet (B). The ribbon-like structures in the pictures are composed of chains of amino acids hydrogen-bonded through their side chains.

Advanced Concepts

Specialized secondary structures can identify functions of proteins. **Zinc finger** motifs are domains frequently found in proteins that bind to DNA.[93] These structures consist of two beta sheets followed by an alpha helix with a stabilizing zinc atom. There are three types of zinc fingers, depending on the arrangement of cysteine residues in the protein sequence. Another example of specialized secondary structure is the **leucine zipper,** also found in transcription factors.[94] The conserved sequence has a leucine or other hydrophobic residue at each seventh position for approximately 30 amino acids arranged in an alpha-helical conformation such that the leucine side chains radiate outwardly to facilitate association with other peptides of similar structure. Because other amino acids besides leucine can participate in this interaction, the term *basic zipper,* or bZip, has been used to describe this type of protein structure. Another similar structure found in transcriptional regulators is the helix-loop helix,[95] consisting of basic amino acids that bind consensus DNA sequences (CANNTG) of target genes. This structure is sometimes confused with the helix-turn helix. The helix-turn helix is two alpha helices connected by a short sequence of amino acids, a structure that can easily fit into the major groove of DNA.

The Sp1 protein, a eukaryotic transcription regulator, contains a zinc finger motif. The side chains of histidine (H) and cysteine (C)—part of the zinc finger amino acid sequence, $\mathrm{C}\text{-}X_{2\text{-}4}\text{-}\mathrm{C}\text{-}X_3\text{-}\mathrm{F}\text{-}X_5\text{-}\mathrm{L}\text{-}X_2\text{-}\mathrm{H}\text{-}X_3\text{-}\mathrm{H}$ (~23 amino acids)—bind a Zn atom in the active protein. Sp1 binds DNA in the regulatory region of genes. Note: For the single-letter amino acid code, see Tables 1.7 and 1.8. *X* in this consensus denotes any amino acid.

The lambda repressor, a transcription factor of the bacteriophage lambda, has a helix-turn-helix motif. One of each of the helices fits into the major groove of the DNA. Lambda repressor prevents transcription of genes necessary for active growth of the bacteriophage, leading to host-cell lysis.

The secondary structures of proteins are further folded and arranged into a **tertiary structure.** Tertiary structure is also important for protein function. If a protein loses its tertiary structure, it is denatured. Mutations in DNA that substitute different amino acids in the primary structure can also alter tertiary structure. Denatured or improperly folded proteins are not functional. Proteins can also be denatured by heat (e.g., the albumin in egg white) or by conformations forced on innocuous peptides by infectious **prions.** Aggregations of prion-induced aberrantly folded proteins cause transmissible spongiform encephalopathies, such as Creutzfeldt–Jakob disease and bovine spongiform encephalitis (mad cow disease).

Two proteins bound together to function form a dimer, three form a trimer, and four form a tetramer. Proteins that work together in this way are called **oligomers,** each component protein being a **monomer.** This is the **quaternary structure** of proteins. The combinatorial nature of protein function may account for the genetic complexity of higher organisms without a concurrent increase in gene number.

Proteins are classified according to function as enzymes and as transport, storage, motility, structural, defense, or regulatory proteins. Enzymes and transport, defense, and regulatory proteins are usually globular in nature, making them soluble and allowing them to diffuse freely across membranes. Structural and motility proteins are fibrous and insoluble.

In contrast to simple proteins that have no other components except amino acids, **conjugated proteins** do have components other than amino acids. The nonprotein component of a conjugated protein is the nonprotein **prosthetic group.** Examples of conjugated proteins are those covalently attached to lipids (lipoproteins), for example, low-density lipoproteins; sugars (glycoproteins), for example, mucin in saliva; and metal atoms (metalloproteins), for example, ferritin. One of the most familiar examples of a conjugated protein is hemoglobin. Hemoglobin is a tetramer, having four Fe^{2+}-containing heme groups, one covalently attached to each monomer.

Genes

A **gene** is defined as the ordered sequence of nucleotides on a **chromosome** that encodes a specific functional product. A gene is the fundamental physical and functional unit of inheritance. The physical definition of a gene was complicated in early studies because of the methods used to define units of genetic inheritance.

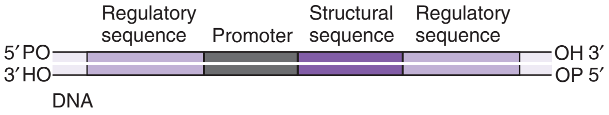


FIGURE 1.37 A gene contains not only structural (coding) sequences but also sequences important for regulated transcription of the gene. These include the promoter, where RNA polymerase binds to begin transcription, and regulatory regions, where transcription factors and other regulatory factors bind to stimulate or inhibit transcription by RNA polymerase.

Genes were first studied by tracking mutations (changes in the order or sequence of nucleotides in the DNA) that took away their function and observing the resulting phenotype. A gene was considered that part of a chromosome responsible for the phenotype affected by mutation. Genes were not delineated well in terms of their physical size but were mapped relative to each other based on the frequency of recombination between them.

A gene contains not only **structural sequences** that code for an amino acid sequence but also **regulatory sequences** that are important for the regulated expression of the gene (Fig. 1.37). Cells expend a good deal of energy to coordinate protein synthesis so that the proper proteins are available at specific times and in specific amounts. Loss of this controlled expression will result in an abnormal phenotype, even though there may be no changes in the structural sequence of the gene. The failure to appreciate the importance of proximal and distal regulatory sequences was another source of confusion in early efforts to define a gene. Regulatory effects and the interaction between proteins still challenge the interpretation of genetic analyses in the clinical laboratory.

The Genetic Code

The nature of a gene was further clarified with the deciphering of the **genetic code** by Francis Crick, Marshall Nirenberg, Philip Leder, Gobind Khorana, and Sydney Brenner.[40,96-99] The genetic code is not information in itself but is a dictionary to translate the 4-nucleotide sequence information in DNA to the 20–amino acid sequence information in proteins.

The triplet nature of the genetic code was surmised based on mathematical considerations. It was reasoned that the smallest set of four possible letters that would yield enough unique groups to denote 20 different amino acids was three. A 1-nucleotide code could only account for $4^1 = 4$ different amino acids, whereas a 2-nucleotide code would yield $4^2 = 16$ different possibilities. A 3-nucleotide code would give $4^3 = 64$ different possibilities, enough to account for all 20 amino acids.

Historical Highlights

In the early 1960s, Seymour Benzer used the T4 bacteriophage to investigate the gene more closely. By mixing the phage with several different phenotypes, he could observe the restoration of normal phenotype (complementation) of one mutant by introducing a phage with a different mutation. Benzer could distinguish mutations that could **complement** each other, even though they affected the same phenotype and mapped to the same location in the phage DNA. He determined that these were mutations in different places within the same gene. He organized many mutants into a series of sites along the linear array of the phage chromosome so that he could structurally define the gene as a continuous linear span of genetic material.

Historical Highlights

The interesting history of the breaking of the genetic code began with a competitive scramble. A physicist and astronomer, George Gamow, organized a group of scientists to concentrate on the problem. They called themselves the RNA Tie Club. Each of the 20 members wore a tie emblazoned with a depiction of RNA and a pin depicting a different amino acid. The group met regularly during the 1950s. They made some progress, including Watson and Crick's "adaptor" hypothesis (predicting tRNA) and Gamow's mathematical prediction of three nucleotides coding for one amino acid. Ultimately, however, they did not exclusively break the genetic code.

The challenge was to decipher this triplet code and prove its function. The simplest way to prove the code would have been to determine the order of nucleotides in a stretch of DNA coding for a protein and compare it with the order of amino acids in the protein. In the early 1960s, protein sequencing was possible, but only limited DNA sequencing was available. Marshall Nirenberg made the initial attempts at the code by using short synthetic pieces of RNA to support protein synthesis in a cell-free extract of *E. coli*. In each of 20 tubes, he placed a different radioactive amino acid, then added cell lysate from *E. coli* and an RNA template. In the first definitive experiment, the input RNA template was a polymer of uracil, UUUUUUU. . . . If the input template supported the synthesis of protein, the radioactive amino acids would be joined together, and the radioactivity would be detected in a precipitable protein. On May 27, 1961, Nirenberg measured radioactive protein levels from 19 of the 20 vials at around 70 counts/mg; however, one vial, which contained the amino acid phenylalanine, yielded protein of 38,000 counts/mg. If the 3-nucleotide code was correct, then UUU was the codon for phenylalanine.

After the first successful demonstration of this strategy, other templates were tested. Each synthetic nucleic acid incorporated different amino acids, based on the composition of bases in the RNA sequence. Codes for phenylalanine (UUU), proline (CCC), lysine (AAA), and glycine (GGG) were soon deduced from the translation of RNA synthesized from a single nucleotide population. More of the code was indirectly deduced using mixtures of nucleotides at different proportions. For instance, an RNA molecule synthesized from a 2:1 mixture of U and C polymerized mostly phenylalanine and leucine into protein. Similar tests with other nucleotide mixtures resulted in distinct amino acid incorporations. Although the nucleotide content of each RNA molecule in these tests was known, the exact order of nucleotides in the triplet was not known, leading to inconclusive results.

Nirenberg and Leder used another technique to get at the basic structure of the code. They observed the binding of specific amino acids to three-base RNA molecules (triplets) in ribosome–tRNA mixtures. By noting which triplet/amino acid combination resulted in binding of the amino acid to ribosomes, they were able to assign 50 of the 64 possible triplets to specific amino acids.

Meanwhile, Gobind Khorana had developed another system. He synthesized longer RNA polymers of known nucleotide sequence. With polynucleotides of repeated sequence, he could predict and then observe the peptides that would come from that RNA. For example, a polymer consisting of two bases, such as . . . UCUCUCUCUCUCUC . . ., was expected to code for a peptide of two different amino acids, one coded for by UCU and one by CUC. This polymer yielded a peptide with the sequence . . . Ser-Leu-Ser-Leu. . . . This experiment did not indicate which triplet coded for which amino acid, but combined with the results from Nirenberg and Leder, the UCU was assigned to serine and CUC to leucine.

By 1965, all 64 triplets, or **codons,** were assigned to amino acids (Fig. 1.38). Once the code was confirmed, specific characteristics of it were apparent. The code is redundant, such that all but two amino acids (methionine and tryptophan) have more than one codon. Triplets coding for the same amino acid are similar, often differing in the third base of the triplet. Crick first referred to this as wobble in the third position.[100] Wobble is also used to describe the movement of the base in the third position of the triplet to form novel pairing between the carrier tRNA and the mRNA template during protein translation. Wobble can also be observed in the third position of the anticodon as well as the codon.[101] Codon selection may be important for the differentiation of human tissues and may also have a role in the development of diseases, such as cancer, where differentiation pathways are altered.[102] Thus, all amino acids except leucine, serine, and arginine are selected by the first two letters of the genetic code. The first two letters, however, do not always specify unique amino acids. For example, CA begins the code for both histidine and glutamine. Three codons—UAG, UAA, and UGA—that terminate protein synthesis are termed **nonsense codons.** UAG, UAA, and UGA were named amber, ocher, and opal, respectively, when they were first defined in bacterial viruses.

The characteristics of the genetic code have consequences for molecular analysis. Mutations or changes in the DNA sequence will have different effects on phenotype, depending on the resultant changes in the amino acid sequence. Accordingly, mutations range from phenotypically silent to drastic. This will be discussed in more detail in later chapters.

First position	Second position of codon				Third position
	U	C	A	G	
U	UUU Phenylalanine	UCU Serine	UAU Tyrosine	UGU Cysteine	U
	UUC	UCC	UAC	UGC	C
	UUA Leucine	UCA	UAA Ter (end)	UGA Ter (end)	A
	UUG	UCG	UAG Ter (end)	UGG Tryptophan	G
C	CUU Leucine	CCU Proline	CAU Histidine	CGU Arginine	U
	CUC	CCC	CAC	CGC	C
	CUA	CCA	CAA Glutamine	CGA	A
	CUG	CCG	CAG	CGG	G
A	AUU Isoleucine	ACU Threonine	AAU Asparagine	AGU Serine	U
	AUC	ACC	AAC	AGC	C
	AUA	ACA	AAA Lysine	AGA Arginine	A
	AUG Methionine	ACG	AAG	AGG	G
G	GUU Valine	GCU Alanine	GAU Aspartic acid	GGU Glycine	U
	GUC	GCC	GAC	GGC	C
	GUA	GCA	GAA Glutamic acid	GGA	A
	GUG	GCG	GAG	GGG	G

FIGURE 1.38 The genetic code. Codons are read as the nucleotide in the left column, then the row at the top, and then the right column. Note how there are up to six codons for a single amino acid. Only methionine and tryptophan have a single codon. Note also the three termination codons (ter): UAA, UAG, UGA.

An interesting observation about the genetic code is that, with limited exceptions, the repertoire of amino acids is limited to 20 in all organisms, regardless of growing environments. Thermophilic and cryophilic organisms adapt to growth at 100°C and freezing temperatures, respectively, not by using structurally different amino acids but by varying the combinations of the naturally occurring amino acids. Cells have strict control and editing systems to protect the genetic code and avoid incorporation of unnatural amino acids into proteins. Studies have shown that it is possible to manipulate the genetic code to incorporate modified amino acids.[103,104] This ability to introduce chemically or physically reactive sites into proteins in vivo has significant implications in biotechnology.

Advanced Concepts

In humans, the termination codon UGA also codes for selenocysteine. Selenoproteins have UGA codons in the middle of their coding regions. In the absence of selenium, protein synthesis stops prematurely in these genes.

TRANSLATION

Amino Acid Charging

After transcription of the sequence information in DNA to RNA, the transcribed sequence must be transferred into proteins. Through the genetic code, a specific nucleic acid sequence is translated to an amino acid sequence and, ultimately, to a phenotype. As proposed in the adaptor hypothesis, there must be a molecular factor that can recognize components of both nucleic acid and protein sequences. This factor is transfer RNA (tRNA). Within the 75- to 95-ribonucleotide sequence of each tRNA is a three-base anticodon, complementary to the codon of a specific amino acid. With the redundancy of the genetic code, there are over 50 tRNAs in humans and 40 in bacteria.[105,106] There are also tRNAs with the same anticodon but with a sequence outside of the anticodon or tRNA **isodecoders,** which are expressed differently in different cells and different stages of development.[107]

Protein synthesis starts with activation of the amino acids by covalent attachment to tRNA, or tRNA **charging,** a reaction catalyzed by 20 **aminoacyl tRNA synthetases.** The Mg^{++}-dependent **charging** reaction was first described by Hoagland and Zamecnik, who observed that amino acids incubated with ATP and the cytosol fraction of liver cells became attached to heat-soluble RNA (tRNA).[108] The reaction takes place in two steps. First, the amino acid is activated by the addition of AMP:

$$\text{amino acid} + \text{ATP} \rightarrow \text{aminoacyl-AMP} + \text{PPi}$$

Second, the activated amino acid is joined to the tRNA:

$$\text{aminoacyl-AMP} + \text{tRNA} \rightarrow \text{aminoacyl-tRNA} + \text{AMP}$$

The product of the reaction is an ester bond between the 3′ hydroxyl of the terminal adenine of the tRNA and the carboxyl group of the amino acid.

Advanced Concepts

According to evolutionary theory, the genetic code has evolved over millions of years of selection. An interesting analysis compared the natural genetic code shared by all living organisms with millions of other possible triplet codes generated by computer (four nucleotides coding for 20 amino acids).[109] The results showed that the natural code was significantly more resistant to damaging changes (mutations in the DNA sequence) compared with the other possible codes. The code is still undergoing "fine-tuning" through common mechanisms in prokaryotes and eukaryotes.[110]

Advanced Concepts

Once the amino acid is esterified to the tRNA, it makes no difference in the specificity of its addition to the protein. The fidelity of translation is now determined by the anticodon of the tRNA as an adaptor between mRNA and the growing protein. This association has been exploited by attaching amino acids synthetically to selected tRNAs. If amino acids are attached to tRNAs carrying anticodons to UAG, UAA, or UGA, the peptide chain will continue to grow instead of terminating at the stop codon. These tRNAs are called suppressor tRNAs because they can suppress point mutations (single-base-pair changes) that generate stop codons within a protein-coding sequence. Suppression of premature termination mutations in mutated genes has been suggested as a type of gene therapy for mutation-driven diseases.[111]

There are 20 amino acid tRNA synthetase enzymes, one for each amino acid. Designated as class I and class II synthetases, these enzymes interact, respectively, with the minor or major groove of the tRNA acceptor arm. Both classes also recognize tRNAs by their anticodon sequences and amino acids by their side chains. Only the appropriate tRNA and amino acid will fit into its cognate synthetase (Fig. 1.39). An errant amino acid bound to the wrong synthetase will dissociate rapidly before any conformation changes and charging can occur. In another level of editing, mischarged aminoacylated tRNAs are hydrolyzed at the point of release from the enzyme.

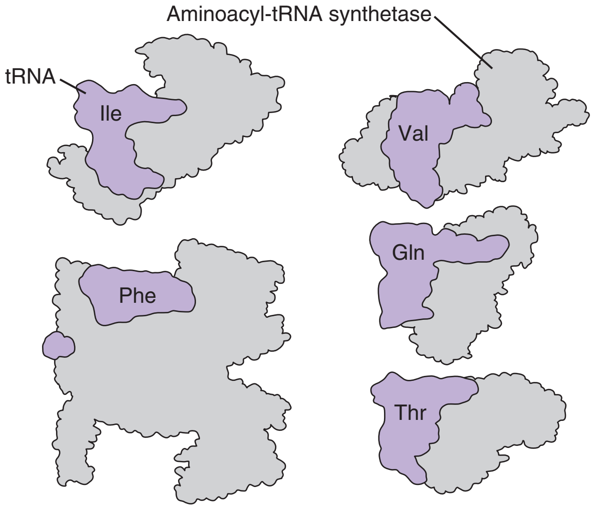


FIGURE 1.39 Five aminoacyl tRNA synthetases. Each enzyme is unique for a tRNA and its matching amino acid. The specificity of these enzymes is key to the fidelity of translation.

Protein Synthesis

Translation takes place on ribosomes, ribonucleoprotein particles first observed by electron microscopy of animal cells. In the early 1950s, Paul Zamecnik demonstrated that these particles were the site of protein synthesis in bacteria.[112] There can be as many as 10 million ribosomes in a eukaryotic cell, depending on cell type, and over 70,000 ribosomes in a bacterial cell, depending on the growth rate of the cell.

Ribosomal structure is similar in prokaryotes and eukaryotes (see Fig. 1.27). In prokaryotes, 70S ribosomes are assembled from a 30S small subunit and a 50S large subunit, in association with mRNA and initiating factors. (S stands for sedimentation units in density-gradient centrifugation, a method used to determine the sizes of proteins and protein complexes.) The 30S subunit (1 million daltons) is composed of a 16S ribosomal RNA (rRNA) and 21 ribosomal proteins. The 50S subunit (1.8 million daltons) is composed of a 5S rRNA, a 23S rRNA, and 34 ribosomal proteins. Eukaryotic ribosomes are slightly larger (80S) and more complex, with a 40S small subunit (1.3 million daltons) and a 60S subunit (2.7 million daltons). The 40S subunit is made up of an 18S rRNA and about 30 ribosomal proteins. The 60S subunit contains a 5S rRNA, a 5.8S rRNA, a 28S rRNA, and about 40 ribosomal proteins.

Protein synthesis in the ribosome almost always starts with the amino acid methionine in eukaryotes and *N*-formylmethionine in bacteria, mitochondria, and chloroplasts. Initiating factors that participate in the formation of the ribosome complex differentiate the initiating methionyl tRNAs from those that add methionine internally to the protein.[113] In protein translation, the small ribosomal subunit first binds to initiation factor 3 (IF-3) and then to specific sequences near the 5′ end of the mRNA, the **ribosomal binding site.** This guides the AUG codon (the “start” codon) to the proper place in the ribosomal subunit. Another initiation factor, IF-2 bound to GTP and the initiating tRNA^{Met} or $\text{tRNA}^{\text{fMet}}$, then joins the complex (Fig. 1.40). The large ribosomal subunit then associates with the hydrolysis of GTP and release of GDP and phosphate, IF-2, and IF-3. The resulting functional 70S or 80S ribosome is the initiation complex. In this complex, the tRNA^{Met} or $\text{tRNA}^{\text{fMet}}$ is situated in the peptidyl site **(P site)** of the functional ribosome. tRNA^{Met} or $\text{tRNA}^{\text{fMet}}$ can only bind to the

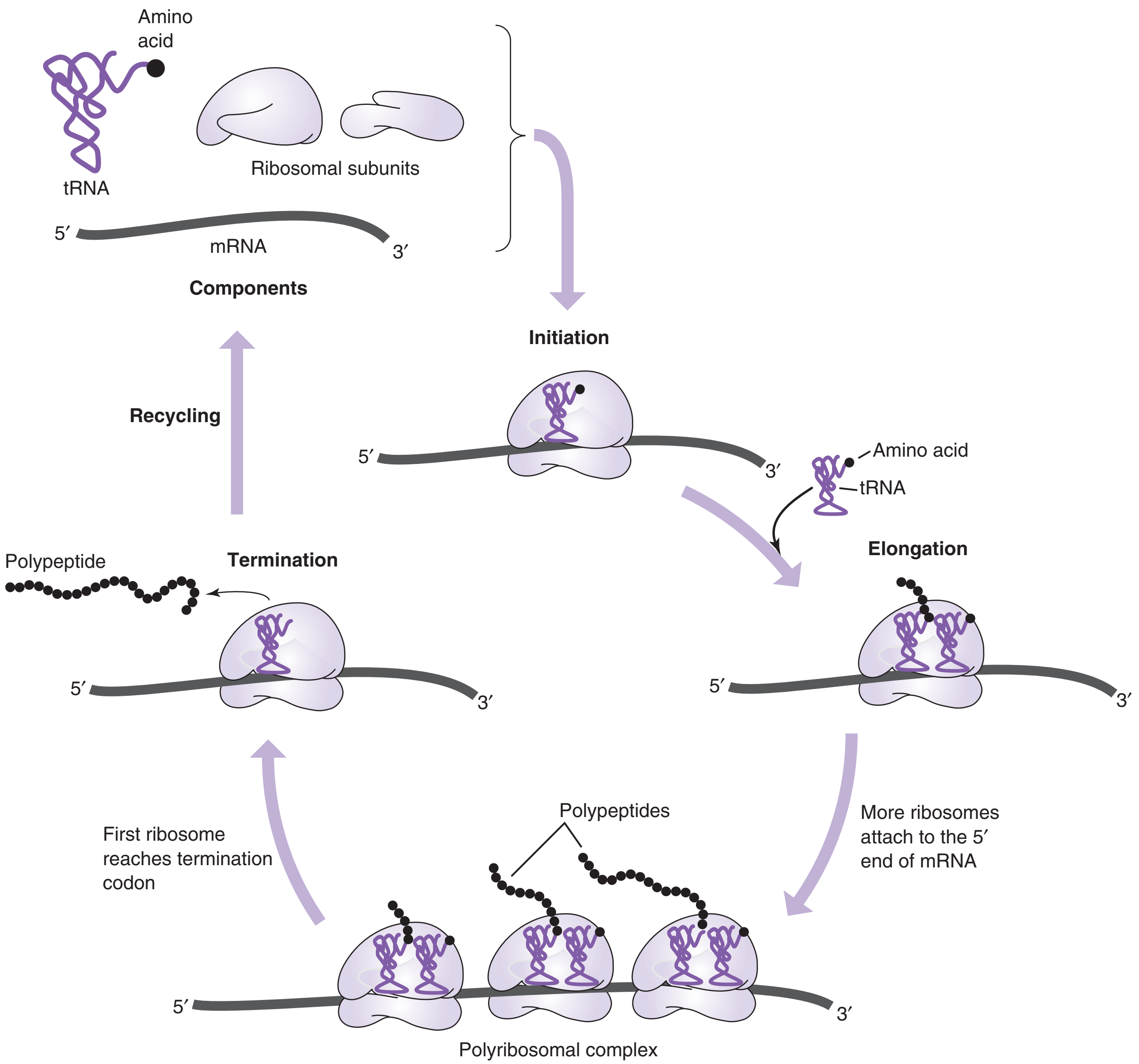

Amino acid
tRNA
Ribosomal subunits
5′
mRNA
3′
Components
Initiation
5′
3′
Amino acid
tRNA
Elongation
5′
3′
Recycling
Polypeptide
Termination
5′
3′
More ribosomes attach to the 5′ end of mRNA
First ribosome reaches termination codon
Polypeptides
5′
3′
Polyribosomal complex

FIGURE 1.40 Assembly of the small ribosome subunit with mRNA and then the large ribosomal subunit; charged tRNA initiates RNA synthesis (initiation). Binding of charged tRNAs and formation of the peptide bond produce the growing polypeptide (elongation). Several ribosomes can simultaneously read a single mRNA (polyribosome complex). When the complex encounters a nonsense codon, protein synthesis stops (termination), and the components are recycled.

P site in the ribosome, which is formed in combination by both ribosomal subunits. In contrast, all other tRNAs bind to an adjacent site, the aminoacyl site **(A site)** of the ribosome.

Synthesis proceeds in the elongation step where the tRNA carrying the next amino acid binds to the A site of the ribosome in a complex with elongation factor Tu (EF-Tu) and GTP (Fig. 1.41). The fit of the incoming tRNA takes place by recognition and then proofreading of the codon–anticodon base pairing. Hydrolysis of GTP by EF-Tu occurs between these two steps.[114] The EF-Tu-GDP is released, and the EF-Tu-GTP is regenerated by another elongation factor, EF-Ts. Although these interactions ensure the accurate pairing of the first two codon positions, the pairing at the third position is not as stringent, which might be related to the wobble in the genetic code.

The first peptide bond is formed between the amino acids in the A and P sites by transfer of the *N*-formylmethionyl group of the first amino acid to the amino group of the second amino acid, generating a dipeptidyl-tRNA in the A site. This step is catalyzed by an enzymatic activity in the large subunit, **peptidyl transferase.** This activity is mediated through RNA, but proteins in the vicinity of the active site of the peptidyl transferase center of the ribosome may influence its organization or the efficiency of the reaction.[115,116] After formation of the peptide bond, the ribosome moves, shifting the dipeptidyl-tRNA from the A site to the P site with the release of the "empty" tRNA from a third position, the **E site,** of the ribosome. This movement (translocation) of tRNAs across a distance of 20 angstroms from the A to the P site and 28 angstroms from the P to the E site requires elongation factor EF-G. As the ribosomal complex moves along the mRNA, the growing peptide chain is always attached to the incoming amino acid. Two GTPs are hydrolyzed to GDP with the addition of each amino acid. This energy-dependent translocation occurs with shifting and rotation of ribosomal subunits (Fig. 1.42).

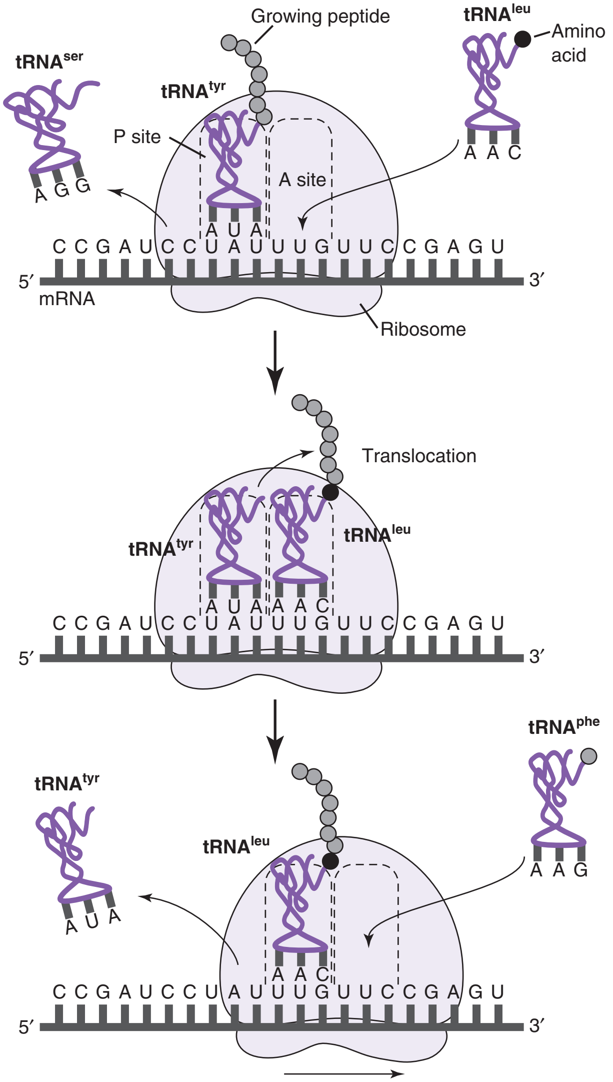


FIGURE 1.41 Incoming charged tRNAs bind to the A site of the ribosome, guided by matching codon–anticodon pairing. After formation of the peptide bond between the incoming amino acid and the growing peptide, the ribosome moves to the next codon in the mRNA, translocating the peptide to the P site and creating another A site for the next tRNA.

During translation, the growing polypeptide begins to fold into its mature conformation. This process is assisted by molecular **chaperones.**[117] These specialized proteins bind to the large ribosomal subunit, forming a hydrophobic pocket that holds the emerging polypeptide (Fig. 1.43). Chaperones apparently protect the growing unfinished polypeptides until they can be safely folded into mature protein. In the absence of this activity, unfinished proteins might bind to each other and form nonfunctional aggregates. Termination of the amino acid chain is signaled by one of the three nonsense, or termination, codons—UAA, UAG, or UGA—which are not charged with an amino acid. When the ribosome

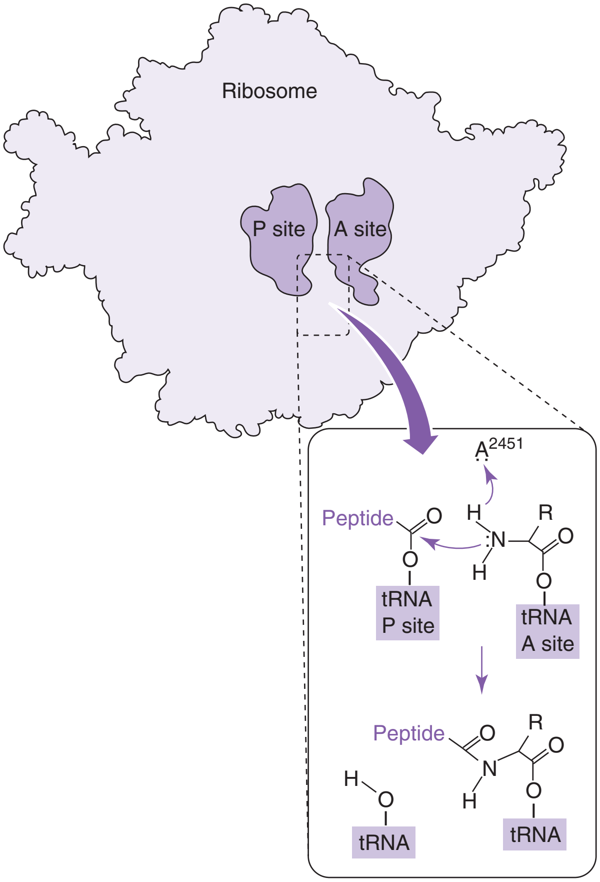

Ribosome
P site
A site
A^{2451}
Peptide
O
H
N
H
R
O
O
tRNA
P site
O
tRNA
A site
Peptide
O
R
O
N
H
H
O
tRNA
O
tRNA

FIGURE 1.42 Protein synthesis (translation) as it takes place in the ribosome. The peptide bond is formed in an area between the large and small subunits of the ribosome. The ribozyme theory holds that the ribosome is an enzyme that functions through RNA and not protein. The close proximity of only RNA to this site is evidence for the ribozyme theory.

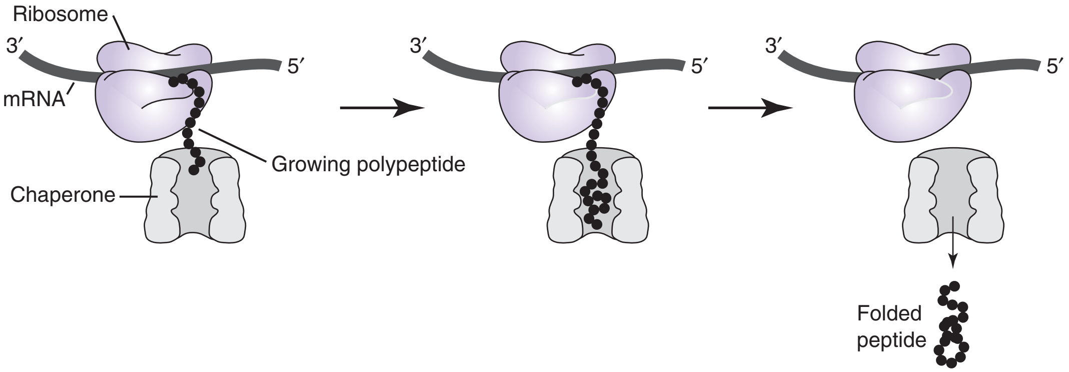

Ribosome
3′
5′
mRNA
Growing polypeptide
Chaperone
3′
5′
3′
5′
Folded peptide

FIGURE 1.43 Molecular chaperones catch the growing peptide as it emerges from the active site. The peptide goes through stages of holding *(left)*, folding *(center)*, and release *(right)*. When the protein is completely synthesized and released from the ribosome, it should be in its folded state. This protects the nascent (growing) peptide from harmful interactions with other proteins in the cell before it has had an opportunity to form its protective and active tertiary structure.

encounters a termination codon, termination, or **release factors** (R_1, R_2, and S in *E. coli*), will cause hydrolysis of the finished polypeptide from the final tRNA, release of that tRNA from the ribosome, and dissociation of the large and small ribosomal subunits. In eukaryotes, termination codon–mediated binding of polypeptide chain release factors (eRF1 and eRF3) triggers hydrolysis of peptidyl-tRNA at the ribosomal peptidyl transferase center.[118,119] *E. coli* can synthesize a 300- to 400-amino acid protein in 10 to 20 seconds. Because the protein takes on its secondary structure as it is being synthesized, it already has its final conformation when it is released from the ribosome.

Advanced Concepts

In eukaryotes, accumulation of unfolded proteins in the endoplasmic reticulum elicits an unfolded protein response (ER stress), intended to correct the problem by cessation of protein synthesis or induction of more chaperones. If correction doesn't occur, ER stress will send the cell into a programmed death (apoptosis). ER stress has been found to be associated with a number of diseases, including diabetes, neurodegenerative disorders, cancer, and cardiovascular disease.[120]

In bacteria, translation and transcription occur simultaneously. In nucleated cells, the majority of translation occurs in the cytoplasm. Several lines of evidence suggest, however, that some translation might also occur in the nucleus. One line of evidence is that nuclei contain factors required for translation. Furthermore, isolated nuclei can aminoacylate tRNAs and incorporate amino acids into proteins.

A cellular surveillance system for RNA with premature termination codons, **nonsense-mediated decay (NMD),**[121] a degradation of messenger RNAs with premature termination codons, supposedly occurred in mammalian nuclei. Further investigations have shown, however, that NMD may not occur in the nuclei of lower eukaryotes.

STUDY QUESTIONS

DNA Structure and Function

1. What is the function of DNA in the cell?
2. Compare the structure of the nitrogen bases. How do purines and pyrimidines differ?
3. Write the complementary sequence to the following:

   5′AGGTCACGTCTAGCTAGCTAGA3′

4. Which of the ribose carbons participate in the phosphodiester bond?
5. Which of the ribose carbons carries the nitrogen base?
6. Why does DNA polymerase require primase activity?

Restriction Enzyme Analysis

1. A plasmid was digested with the enzyme *Hpa*II. On agarose gel electrophoresis, you observe three bands, 100, 230, and 500 bp.
   a. How many *Hpa*II sites are present in this plasmid?
   b. What are the distances between each site?
   c. What is the size of the plasmid?
   d. Draw a picture of the plasmid with the *Hpa*II sites.

   A second cut of the plasmid with *Bam*H1 yields two pieces, 80 and *x* bp.

   e. How many *Bam*H1 sites are in the plasmid?
   f. What is *x* in base pairs (bp)?
2. How would you determine where the *Bam*H1 sites are in relation to the *Hpa*II sites?
3. The plasmid has one *Eco*R1 site into which you want to clone a blunt-ended fragment. What type

of enzyme could turn an *Eco*R1 sticky end into a blunt end?

Recombination and DNA Transfer

1. Describe how DNA moves from cell to cell by (a) conjugation, (b) transduction, and (c) transformation.
2. Which of the three interactions in question 1 would be prevented by a barrier between two mating strains that stops bacterial cells, but not smaller particles?
3. After meiosis, gametes produced from diploid organisms are ___________ (haploid/diploid).

RNA Secondary Structure

1. Draw the secondary structure of the following RNA. The complementary sequences (inverted repeat) are underlined.

   5′CAUGUUCAGCUCAUGUGAACGCU3′

2. Underline *two* inverted repeats in the following RNA.

5′CUGAACUUCAGUCAAGCAAAGAGUUUGCACUG3′

RNA Processing

1. Name three processing steps undergone by messenger RNA.
2. What is the function of polyadenylate polymerase?
3. What is unusual about the phosphodiester bond formed between mRNA and its 5′ cap?
4. The parts of the hnRNA that are translated into protein (open reading frames) are the __________.
5. The intervening sequences that interrupt protein-coding sequences in hnRNA are _____________.

Proteins and the Genetic Code

1. Indicate whether the following peptides are hydrophilic or hydrophobic.
   - a. MLWILAA
   - b. VAIKVLIL
   - c. CSKEGCPN
   - d. SSIQKNET
   - e. YAQKFQGRT
   - f. AAPLIWWA
   - g. SLKSSTGGQ
2. Is the following peptide positively or negatively charged at neutral pH?

   GWWMNKCHAGHLNGVYYQGGTY

3. Consider an RNA template made from a 2:1 mixture of C:A. What would be the three amino acids *most* frequently incorporated into protein?
4. What is the peptide sequence encoded in AUAUAUAUAUAUAUA . . .?
5. Write the *anticodons* 5′ to 3′ of the following amino acids:
   - a. L
   - b. T
   - c. M
   - d. H
   - e. R
   - f. I
6. A protein contains the sequence LGEKKWCLRVNPKGLDESKDYLSLKSKYLLL. What is the likely function of this protein? (Note: See Advanced Concepts box.)
7. A histone-like protein contains the sequence PKKGSKKAVTKVQKKDGKKRKRSRK. What characteristic of this sequence makes it likely to associate with DNA?
8. A procedure for digestion of DNA with a restriction enzyme includes a final incubation step

of 5 minutes at 95°C. What is the likely purpose of this final step?

9. What is a ribozyme?

10. Name the nonprotein prosthetic groups for the following conjugated proteins:
glycoprotein
lipoprotein
metalloprotein

References

1. Watson J. *Molecular biology of the gene*. San Francisco, CA: W. A. Benjamin, 1965.
2. Mirsky AE. The discovery of DNA. *Scientific American* 1968;218: 78–88.
3. Watson J, Crick FH. DNA structure. *Nature* 1953;171:737.
4. Watson J, Hopkins N, Roberts J, Steitz J, Weiner A. *The molecular biology of the gene*. 4th ed. Redwood City, CA: Benjamin/ Cummings, 1987.
5. Eschenmoser A. Chemical etiology of nucleic acid structure. *Science* 1999;284:2118–2124.
6. Egholm M, Buchardt O, Christensen L, Behrens C, Freier SM, Driver DA, Berg RH, Kim SK, Norden B, Nielsen PE. PNA hybridizes to complementary oligonucleotides obeying the Watson–Crick hydrogen-bonding rules. *Nature* 1993;365:566–568.
7. Demidev VV, Yavnilovich MV, Belotserkovskii BP, Frank-Kamenetskii MD, Nielsen P. Kinetics and mechanism of polyamide ("peptide") nucleic acid binding to duplex DNA. *Proceedings of the National Academy of Sciences* 1995;92:2637–2641.
8. Ray A, Norden B. Peptide nucleic acid (PNA): its medical and biotechnical applications and promise for the future. *Federation of American Societies for Experimental Biology Journal* 2000;14:1041–1060.
9. Hall L, Le Febre KM, Deml SM, Wohlfiel SL, Wengenack NL. Evaluation of the yeast Traffic Light PNA FISH probes for identification of *Candida* species from positive blood cultures. *Journal of Clinical Microbiology* 2012;50:1446–1448.
10. Dean D. Peptide nucleic acids: versatile tools for gene therapy strategies. *Advanced Drug Delivery Reviews* 2000;44:81–95.
11. Chargaff E. Chemical specificity of nucleic acids and mechanisms of their enzymatic degradation. *Experimentia* 1950;6:201–209.
12. Meselson M, Stahl FW. The replication of DNA in *Escherichia coli*. *Proceedings of the National Academy of Sciences* 1958;44:671–682.
13. Okazaki R, Okazaki T, Sakabe K, Sugimoto K, Kainuma R, Sugino R, Zwatsuki N. In vivo mechanism of DNA chain growth. *Cold Spring Harbor Symposium on Quantitative Biology* 1968;33:129–144.
14. Kornberg A. The synthesis of DNA. *Scientific American* 1968.
15. Kutchta R, Gudrun S. Mechanism and evolution of DNA primases. *Biochimica Biophysica Acta* 2010;1804:1180–1189.
16. Pandey M, Syed S, Donmez I, Patel G, Ha T, Patel SS. Coordinating DNA replication by means of priming loop and differential synthesis rate. *Nature* 2009;462:940–943.
17. Dervyn E, Suski C, Daniel R, Bruand C, Chapuis J, Errington F, Janniere L, Ehrlich SD. Two essential DNA polymerases at the bacterial replication fork. *Science* 2001;294:1716–1719.
18. Schachman HK, Adler J, Radding CM, Lehman IR, Kornberg A. Enzymatic synthesis of deoxyribonucleic acid. VII. Synthesis of a polymer of deoxyadenylate and deoxythymidylate. *Journal of Biological Chemistry* 1960;235:3242–3249.
19. Ito J, Braithwaite DK. Compilation and alignment of DNA polymerase sequences. *Nucleic Acids Research* 1991;19:4045–4057.
20. Braithwaite DK, Ito J. Complication, alignment, and phylogenetic relationships of DNA polymerases. *Nucleic Acids Research* 1993;21:787–802.
21. Hamilton SC, Farchaus JW, Davis MC. DNA polymerases as engines for biotechnology. *BioTechniques* 2001;31:370–383.
22. Butterer A, Pernstich C, Smith RM, Sobott F, Szczelkun MD, Tóth J. Type III restriction endonucleases are heterotrimeric: comprising one helicase–nuclease subunit and a dimeric methyltransferase that binds only one specific DNA. *Nucleic Acids Research* 2014;42:5139–5150.
23. Sears A, Peakman LJ, Wilson GG, Szczelkun MD. Characterization of the type III restriction endonuclease PstII from *Providencia stuartii*. *Nucleic Acids Research* 2005;33:4775–4787.
24. Jurenaite-Urbanaviciene S, Kazlauskiene R, Urbelyte V, Maneliene Z, Petruyte M, Lubys A, Janulaitisa A, Jurenaite-Urbanaviciene S, Kazlauskiene R, Urbelyte V, Maneliene Z, Petruyte M, Lubys A, Janulaitisa A. Characterization of BseMII, a new type IV restriction–modification system, which recognizes the pentanucleotide sequence 5′-CTCAG(N)10/8. *Nucleic Acids Research* 2001;29:895–903.
25. Wright AV, Nuñez JK, Doudna JA. Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. *Cell* 2016;164:29–44.
26. Marakova K, Koonin EV. Annotation and classification of CRISPR-Cas systems. *Methods in Molecular Biology* 2015; 1311:47–75.
27. Lehman IR. DNA ligase: structure, mechanism, and function. *Science* 1974;186:790–797.
28. Lehman IR. *The enzymes*. New York, NY: Academic Press, 1971.
29. Sgaramella V, Van de Sande JH, Khorana HG. Studies on polynucleotides, C. A novel joining reaction catalyzed by the T4-polynucleotide ligase. *Proceedings of the National Academy of Sciences* 1971;67:1468–1475.
30. Berk AJ, Sharp PA. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. *Cell* 1977;12:721–732.
31. Kornberg RD. Structure of chromatin. *Annual Review of Biochemistry* 1977;46:931–954.
32. Pommier Y, Sun Y, Huang SN, Nitiss JL. Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. *Nature Review Molecular and Cellular Biology* 2016;17: 703–721.
33. Wang J. Interaction between DNA and an *Escherichia coli* protein omega. *Journal of Molecular Biology* 1971;55:523–533.

34. Gellert M, Mizuuchi K, O'Dea MH, Nash HA. DNA gyrase: an enzyme that introduces superhelical turns into DNA. *Proceedings of the National Academy of Sciences* 1976;73:3872–3876.

35. Mendel G. Experiments in plant hybridization. In Peters JA, ed., *Classic papers in genetics*. Upper Saddle River, NJ: Prentice-Hall, 1959 (1865).

36. Beadle GW. Genes and the chemistry of organism. *American Scientist* 1946;34:31–53.

37. Lederberg J, Tatum EL. Sex in bacteria; genetic studies, 1945–1952. *Science* 1953;118:169–175.

38. Jacob F, Wollman E. Viruses and genes. *Scientific American* 1961;204:92–110.

39. Hershey AD, Chase M. Independent function of viral protein and nucleic acid in growth of bacteriophage. *Journal of General Physiology* 1952;26:36–56.

40. Benzer S. The fine structure of the gene. *Scientific American* 1962;206:70–84.

41. Griffith F. Significance of pneumococcal types. *Journal of Hygiene* 1928;27:113–159.

42. Avery O, MacLeod CM, McCarty M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types. I. Induction of transformation by a DNA fraction isolated from pneumococcal type III. *Journal of Experimental Medicine* 1944;79:137–158.

43. Hotchkiss R. The genetic chemistry of the *Pneumococcal* transformations. The *Harvey Lectures*. Cambridge, MA: Academic Press, 1955. pp. 124–144.

44. Clowes R. The molecule of infectious drug resistance. *Scientific American* 1973;228:19–27.

45. Watanabe T, Fukasawa T. Episome mediated transfer of drug resistance in Enterobacteriaceae. II. Elimination of resistance factors with acridine dyes. *Journal of Bacteriology* 1961;81:679–683.

46. Watanabe T. Episome mediated transfer of drug resistance in Enterobacteriaceae. VI. High frequency resistance transfer system in *E. coli*. *Journal of Bacteriology* 1963;85:788–794.

47. Moller JK, Bak AL, Christiansen G, Stenderup A. Extrachromosomal DNA in R factor harboring Enterobacteriaceae. *Journal of Bacteriology* 1976; 125:398–403.

48. Werner M, Thuriaux P, Soutourina J. Structure-function analysis of RNA polymerases I and III. *Current Opinion in Structural Biology* 2009;19:740–745.

49. Tollervey D. Termination by torpedo. *Nature* 2004;432:456–457.

50. Epstein M, Duren WL, Boehnke M. Improved inference of relationship for pairs of individuals. *American Journal of Human Genetics* 2000;67:1219–1231.

51. Kuhn A, Bartsch I, Grummt I. Specific interaction of the murine transcription termination factor TTF I with class-I RNA polymerases. *Nature* 1990;344:559–562.

52. Weinberg RA, Penman S. Processing of 45S nucleolar RNA. *Journal of Molecular Biology* 1970;47:169–178.

53. Ogle JM, Brodersen DE, Clemons WM Jr., Tarry MJ, Carter AP, Ramakrishman V. Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* 2001;292:897–902.

54. Brown A, Shao S, Murray J, Hegde RS, Ramakrishnan V. Structural basis for stop codon recognition in eukaryotes. *Nature* 2015;524:493–496.

55. Edmonds M, Caramela MG. The isolation and characterization of AMP rich polynucleotide synthesized by Ehrlich ascites cells. *Journal of Biological Chemistry* 1969;244:1314–1324.

56. Wickens M, Gonzalez TN. Knives, accomplices and RNA. *Science* 2004;306:1299–1300.

57. Paushkin S, Patel M, Furia BS, Peltz SW, Trotta CR. Identification of a human endonuclease complex reveals a link between tRNA splicing and pre-mRNA 3′ end formation. *Cell* 2004;117:311–321.

58. Tilghman SM, Tiemeier DC, Seidman JG, Peterlin BM, Sullivan M, Maizel JV, Leder P. Intervening sequence of DNA identified in the structural portion of a mouse b-globin gene. *Proceedings of the National Academy of Sciences* 1978;75:725–729.

59. White R, Hogness DS. R loop mapping of the 18S and 28S sequences in the long and short repeating units of *D. melanogaster* rDNA. *Cell* 1977;10:177–192.

60. Perry R, Kelley DE. Existence of methylated mRNA in mouse L cells. *Cell* 1974;1:37–42.

61. Milcarek C, Price R, Penman S. The metabolism of a poly(A) minus mRNA fraction in HeLa cells. *Cell* 1974;3:1–10.

62. Curinha A, Oliveira Braz S, Pereira-Castro I, Cruz A, Moreira A. Implications of polyadenylation in health and disease. *Nucleus* 2014;5:508–519.

63. Pyle A. Group II intron self-splicing. *Annual Review of Biophysics* 2016;45:183–205.

64. Waring R, Davies RW. Assessment of a model for intron RNA secondary structure relevant to RNA self-splicing—a review. *Gene* 1984;28:277–291.

65. Lewin B. *Gene expression 2: Eucaryotic chromosomes*. 2nd ed. New York, NY: John Wiley & Sons, 1980.

66. Both GW, Furuichi Y, Muthukrishman S, Shatkin AJ. Effect of 5′ terminal structure and base composition on polyribonucleotide binding to ribosomes. *Journal of Molecular Biology* 1975;104:637–658.

67. Amara SG, Jonas V, Rosenfield MG, Ory ES, Evans RM. Alternative RNA processing in calcitonin gene expression generates mRNA encoding different polypeptide products. *Nature* 1982;298:240–244.

68. Toro I, Thore S, Mayer C, Basquin J, Seraphin B, Suck D. RNA binding in an Sm core domain: x-ray structure and functional analysis of an archaeal Sm protein complex. *EMBO Journal* 2001;20:2293–2303.

69. Mura C, Cascio D, Sawaya MR, Eisenberg DS. The crystal structure of a heptameric archaeal Sm protein: implications for the eukaryotic snRNP core. *Proceedings of the National Academy of Sciences* 2001;98:5532–5537.

70. Holley R. *The nucleotide sequence of a nucleic acid*. San Francisco, CA: Freeman, 1968.

71. Travers A, Burgess RR. Cyclic reuse of the RNA polymerase sigma factor. *Nature* 1969;222:537–540.

72. Burgess R, Travers AA, Dunn JJ, Bautz EK. Factor stimulating transcription by RNA polymerase. *Nature* 1969;221:43–47.

73. Roeder R, Rutter WJ. Specific nucleolar and nucleoplasmic RNA polymerases. *Proceedings of the National Academy of Sciences* 1970;65:675–682.

74. Jacob S, Sajdel EM, Munro HN. Different responses of soluble whole nuclear RNA polymerase and soluble nucleolar RNA polymerase to divalent cations and to inhibition by a-amanitin. *Biochemical Biophysical Research Communications* 1970;38: 765–770.

75. Lindell T, Weinber F, Morris PW, Roeder RG, Rutter WJ. Specific inhibition of nuclear RNA polymerase by amanitin. *Science* 1970;170:447–448.

76. Qian X, Hamid FM, El Sahili A, Darwis DA, Wong YH, Bhushan S, Makeyev EV, Lescar J. Functional evolution in orthologous cell-encoded RNA-dependent RNA polymerases. *Journal of Biological Chemistry* 2016;291:9295–9309.

77. Hercík K, Kozak J, Šála M, Dejmek M, Hřebabecký H, Zborníková E, Smola M, Ruzek D, Nencka R, Boura E. Adenosine triphosphate analogs can efficiently inhibit the Zika virus RNA-dependent RNA polymerase. *Antiviral Research* 2017;137:131–133.

78. Dickson K, Thompson SR, Gray NK, Wickens M. Poly(A) polymerase and the regulation of cytoplasmic polyadenylation. *Journal of Biological Chemistry* 2001;276:41810–41816.

79. Ryan K, Calvo O, Manley JL. Evidence that polyadenylation factor CPSF-73 is the mRNA 3′ processing endonuclease. *RNA* 2004;10:565–573.

80. Misra A, Green MR. From polyadenylation to splicing: dual role for mRNA 3′ end formation factors. *RNA Biology* 2016;13:259–264.

81. Harder J, Schroder JM. RNase 7, a novel innate immune defense antimicrobial protein of healthy human skin. *Journal of Biological Chemistry* 2002;277:46779–46784.

82. Probst J, Brechtel S, Scheel B, Hoerr I, Jung G, Rammensee HG, Pascolo S. Characterization of the ribonuclease activity on the skin surface. *Genetic Vaccines and Therapy* 2006;4: 1–9.

83. Ponomarenko EA, Poverennaya EV, Ilgisonis EV, Pyatnitskiy MA, Kopylov AT, Zgoda VG, Lisitsa AV, Archakov AI. The size of the human proteome: The width and depth. *International Journal of Analytical Chemistry* 2016;2016:7436849.

84. Schoenmakers E, Carlson B, Agostini M, Moran C, Rajanayagam O, Bochukova E, Tobe R, Peat R, Gevers E, Muntoni F, Guicheney P, Schoenmakers N, Farooqi S, Lyons G, Hatfield D, Chatterjee K. Mutation in human selenocysteine transfer RNA selectively disrupts selenoprotein synthesis. *Journal of Clinical Investigation* 2016;126:992–996.

85. Sanger F, Tuppy H. The amino-acid sequence in the phenylalanyl chain of insulin. I. The identification of lower peptides from partial hydrolysates. *Biochemical Journal* 1951;49: 463–481.

86. Sanger F, Tuppy H. The amino-acid sequence in the phenylalanyl chain of insulin. 2. The investigation of peptides from enzymic hydrolysates. *Biochemical Journal* 1951;49:481–490.

87. Hilterbrand A, Saelens J, Putonti C. CBDB: the codon bias database. *BMC Bioinformatics* 2012;13:62.

88. Zaneveld J, Hamady M, Sueoka N, Knight R. CodonExplorer: an interactive online database for the analysis of codon usage and sequence composition. *Methods in Molecular Biology* 2009;537:207–232.

89. Pauling L, Corey RB. The pleated sheet, a new layer configuration of polypeptide chains. *Proceedings of the National Academy of Sciences* 1951;37:251–256.

90. Pauling L, Corey RB. The structure of synthetic polypeptides. *Proceedings of the National Academy of Sciences* 1951;37: 241–250.

91. Pauling L, Corey RB. The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proceedings of the National Academy of Sciences* 1951;205:205–211.

92. Pauling L, Corey RB. Atomic coordinates and structure factors for two helical configurations of polypeptide chains. *Proceedings of the National Academy of Sciences* 1951;37:235–240.

93. Eom KS, Cheong JS, Lee SJ. Structural analyses of zinc finger domains for specific interactions with DNA. *Journal of Microbiology and Biotechnology* 2016;26:2019–2029.

94. Landschulz W, Johnson PF, McKnight SL. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* 1988;240:1759–1764.

95. Ferre-D’Amare A, Prendergast GC, Ziff EB, Burley SK. Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. *Nature* 1993;363:38–45.

96. Benzer S. On the topography of genetic fine structure. *Proceedings of the National Academy of Sciences* 1961;47: 403–415.

97. Crick F, Barnett L, Brenner S, Watts-Tobin RJ. General nature of the genetic code for proteins. *Nature* 1961;192:1227–1232.

98. Nirenberg M, Leder P, Bernfield M, Brimacombe R, Trupin J, Rottman F, O’Neal C. RNA codewords and protein synthesis, VII: On the general nature of the RNA code. *Proceedings of the National Academy of Sciences* 1965;53:1161–1168.

99. Jones D, Nishimura S, Khorana HG. Studies on polynucleotides, LVI: further synthesis, in vitro, of copolypeptides containing two amino acids in alternating sequence dependent upon DNA-like polymers containing two nucleotides in alternating sequence. *Journal of Molecular Biology* 1966;16:454–472.

100. Crick F, Barnett L, Brenner S, Watts-Tobin RJ. Codon-anticodon pairing: the wobble hypothesis. *Journal of Molecular Biology* 1966;19:548.

101. Das G, Lyngdoh RH. Role of wobble base pair geometry for codon degeneracy: purine-type bases at the anticodon wobble position. *Journal of Molecular Modeling* 2012;18:3805–3820.

102. Hajjari M, Sadeghi I, Salavaty A, Nasiri H, Birgani MT. Tissue specific expression levels of apoptosis involved genes have correlations with codon and amino acid usage. *Genomics and Informatics* 2016;14:234–240.

103. Wang L, Brock A, Herberich B, Schultz PG. Expanding the genetic code of *Escherichia coli*. *Science* 2001;292:498–500.

104. Doring V, Mootz HD, Nangle LA, Hendrickson TL, de Crecy-Lagard V, Schimmel P, Marliere P. Enlarging the amino acid set of *Escherichia coli* by infiltration of the valine coding pathway. *Science* 2001;292:501–504.

105. Iben J, Maraia1 RJ. tRNA gene copy number variation in humans. *Gene* 2014;536:376–384.

106. Deutscher M. Twenty years of bacterial RNases and RNA processing: how we’ve matured. *RNA* 2015;21:597-600.

107. Parisien M, Wang X, Pan T. Diversity of human tRNA genes from the 1000-genomes project. *RNA Biology* 2013;10:1853–1867.

108. Hoagland MB, Keller EB, Zamecnik PC. Enzymatic carboxyl activation of amino acids. *Journal of Biological Chemistry* 1956;218:345–358.

109. Vogel G. Tracking the history of the genetic code. *Science* 1998;281:329–331.

110. Grosjean H, de Crecy-Lagard V, Marck C. Deciphering synonymous codons in the three domains of life: co-evolution with specific tRNA modification enzymes. *FEBS Letters* 2010;584: 252–264.

111. Wang X, Gregory-Evans CY. Nonsense suppression therapies in ocular genetic diseases. *Cellular and Molecular Life Sciences* 2015;72:1931–1938.

112. Zamecnik PC. An historical account of protein synthesis with current overtones—a personalized view. *Cold Spring Harbor Symposium on Quantitative Biology* 1969;34:1–16.

113. Kolitz S, Lorsch JR. Eukaryotic initiator tRNA: Finely tuned and ready for action. *FEBS Letters* 2010;584:396–404.

114. Ruusala T, Ehrenberg M, Kurland CG. Is there proofreading during polypeptide synthesis? *EMBO Journal* 1982;1:741–745.

115. Steitz T, Moore PB. RNA, the first macromolecular catalyst: the ribosome is a ribozyme. *Trends in Biochemical Sciences* 2003;28:411–418.

116. Maracci C, Wohlgemuth I, Rodnina MV. Activities of the peptidyl transferase center of ribosomes lacking protein L27. *RNA* 2015;21:2047–2052.

117. Horwich A. Sight at the end of the tunnel. *Nature* 2004;431: 520–522.

118. Zhouravleva G, Frolova L, Le Goff X, Le Guellec R, Inge-Vechtomov S, Kisselev L, Philippe M. Termination of translation in eukaryotes is governed by two interacting polypeptide chain release factors, eRF1 and eRF3. *EMBO Journal* 1995;14:4065–4072.

119. Frolova L, Merkulova TI, Kisselev LL. Translation termination in eukaryotes: polypeptide release factor eRF1 is composed of functionally and structurally distinct domains. *RNA* 2000;6:381–390.

120. Hotamisligil G, Davis RJ. *Cell signaling and stress responses*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 2016.

121. Lejeune F. Nonsense-mediated mRNA decay at the crossroads of many cellular pathways. *BMB Reports* 2017;50(4):175–185.

Chapter 2

Gene Expression and Epigenetics

Outline

TRANSCRIPTION
- *Transcription Initiation*
- *Transcription Elongation*
- *Transcription Termination*

REGULATION OF TRANSCRIPTION
- *Regulation of Messenger RNA Synthesis at Initiation*
- *Post-Transcriptional Regulation*
- *Post-Translational Regulation*

EPIGENETICS
- *Histone Modification*
- *Nucleic Acid Methylation*

CLASSIFICATION OF EPIGENETIC FACTORS

NONCODING RNAs
- *MicroRNAs*
- *Small Interfering RNAs*
- *Other Small RNAs*
- *Long Noncoding RNAs*

Objectives

2.1 Define *epigenetics,* and list examples of epigenetic phenomena.
2.2 Demonstrate gene regulation using the Lac operon as an example.
2.3 Categorize non-coding RNA and its mechanism of action.
2.4 Describe histone modification and the histone code.
2.5 Explain nucleic acid methylation and its effect on gene expression.
2.6 Identify post-translational protein modifications.

TRANSCRIPTION

DNA can only store information. In order for this information to be utilized, it must be transcribed and then translated into protein, a process called **gene expression.** A specific type of RNA, **messenger RNA (mRNA).** carries the information in DNA to the ribosomes, where it is translated into protein.

Transcription is the copying of one strand of DNA into RNA by a process similar to that of DNA replication. This activity, catalyzed by **RNA polymerase,** occurs mostly in interphase. A single type of RNA polymerase catalyzes the synthesis of all RNA in most prokaryotes. There are three types of RNA polymerases in eukaryotes: RNA polymerase pol I, pol II, and pol III. Pol I and III synthesize noncoding RNA.[1] Pol II is responsible for the synthesis of mRNA, the type of RNA that carries genetic information to be translated into protein (Table 2.1). Evidence suggests that transcription takes place at discrete stations of the nucleus into which the DNA molecules move.[2,3] One of these sites, the nucleolus, is the location of ribosomal RNA synthesis (Table 2.2).

Cells with the same DNA genome (genotype) can have vastly different morphologies and activities (phenotype). Heritable changes in phenotype arise from structural and chemical sequence changes in DNA. More immediate changes in phenotype are brought about by selective expression (transcription and translation) of genes. Thus, genetics and epigenetics combine to adapt cells and organisms to their environment.

TABLE 2.1 RNA Polymerases

Enzyme	Template	Product
E. coli RNA polymerase II	DNA	mRNA
RNA polymerase I	DNA	rRNA
RNA polymerase II	DNA	mRNA
RNA polymerase III	DNA	tRNA, snRNA
Mitochondrial RNA polymerase	DNA	mRNA
Mammalian DNA polymerase α	DNA	primers
HCV RNA polymerase	RNA	Viral genome
Dengue virus RNA polymerase	RNA	Viral genome
PolyA polymerase	None	PolyA tails

Transcription Initiation

RNA polymerase and its supporting accessory proteins assemble at a start site in DNA, a specific sequence of bases called the **promoter.** Sites for initiation of transcription (RNA synthesis) greatly outnumber DNA initiation sites (origins of replication) in both prokaryotes and eukaryotes. There are also many more molecules of RNA polymerase than DNA polymerase in the cell. Although functionally catalyzing the same reaction, RNA polymerases in prokaryotes and eukaryotes differ and work with different supporting proteins to find and bind to DNA in preparation for transcription. In prokaryotes, a basal transcription complex comprised of the large and small subunits of RNA polymerase and additional sigma factors assembles at the start site. The eukaryotic transcription complex requires RNA polymerase and up to 20 additional factors for accurate initiation. Initiation of RNA synthesis is regulated in all organisms so that genes are transcribed as required by specific cell types.

Transcription Elongation

RNA polymerases in both eukaryotes and prokaryotes synthesize RNA using the base sequence of one strand

TABLE 2.2 Cellular Location and Activity of RNA Pol I, II, and III in Eukaryotes

Type	Location	Products	α-Amanitin
I	Nucleolus	18s, 5.8s, 28s rRNA	Insensitive
II	Nucleus	mRNA, snRNA	Inhibited
III	Nucleus	tRNA, 5s rRNA	Inhibited by high concentration

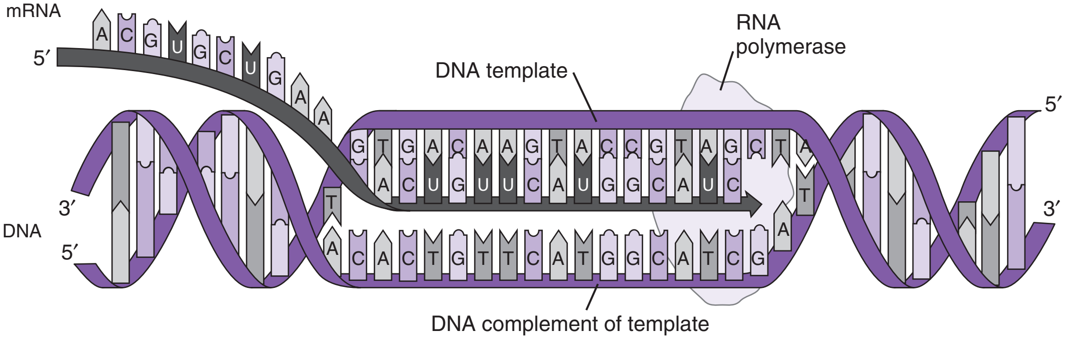

mRNA
5′
A C G U G C U G A A
RNA polymerase
DNA template
5′
G T G A C A A G T A C C G T A G C T A
A C U G U U C A U G G C A U C
3′
DNA
5′
T
A C A C T G T T C A T G G C A T C G
A
T
3′
DNA complement of template

FIGURE 2.1 RNA transcription proceeds by synthesis of the RNA molecule using one strand of the DNA template. The copied strand is complementary to the RNA product, whereas the homologous strand has the same sequence as the RNA product.

of the double helix (the **antisense strand**) as a guide (Fig. 2.1). The **sense strand** of the DNA template has a sequence identical to that of the RNA product (except for the U for T substitution in RNA), but it does not serve as the template for the RNA.

RNA polymerases work more slowly than DNA polymerases (50 to 100 bases/sec for RNA synthesis vs. 1,000 bases/sec for DNA replication) and with less fidelity. The DNA double helix is locally unwound into single strands to allow the assembly and passage of the transcription machinery, forming a transcription bubble.

Unlike DNA synthesis, RNA synthesis does not require priming. The first ribonucleoside triphosphate retains all of its phosphate groups as the RNA is polymerized in the 5′ to 3′ direction. Subsequent ribonucleoside triphosphates retain only the **alpha phosphate,** the one closest to the ribose sugar. The other two phosphate groups are released as orthophosphate during the synthesis reaction.

After initiation of mRNA synthesis, the 5′ end of the growing transcript is covalently attached to a methylated guanine residue (7-methylguanosine) in a 5′ to 5′ covalent bond. Capping occurs in three steps: (1) hydrolysis of the 5′ phosphate from the pre-mRNA, (2) transfer of a guanine monophosphate nucleoside to the 5′ diphosphate end, and (3) methylation of the attached guanine in the N7 position. The cap protects the end of the growing pre-mRNA (or hnRNA) and facilitates ribosome binding and efficient translation.

Transcription Termination

RNA synthesis terminates differently in prokaryotes and eukaryotes. In prokaryotes, RNA synthesis is responsive to protein products, such that high levels of a gene product induce termination of its own synthesis. This synthesis is accomplished in some genes by interactions between RNA polymerase and termination signals in the DNA template. In other genes, an additional transcription complex factor, rho, is required for termination. Rho is a helicase enzyme that associates with RNA polymerase and inactivates the elongation complex at a cytosine-rich termination site in the DNA.[4,5] Rho-independent termination occurs at G:C-rich regions in the DNA, followed by A:T-rich regions. The G:C bases are transcribed into RNA and fold into a short double-stranded **hairpin,** which slows the elongation complex. The elongation complex then dissociates as it reaches the A:T-rich area.

In eukaryotes, mRNA synthesis, catalyzed by pol II, proceeds along the DNA template until a **polyadenylation signal** (polyA site) is encountered. At this point, the process of termination of transcription is activated. There is no consensus sequence in DNA that specifies the termination of transcription. As the polymerase proceeds past the polyA site, the nascent mRNA is released by an endonuclease associated with the carboxy terminal end of pol II. RNA synthesized beyond the site trails out of the polymerase and is bound by another exonuclease that begins to degrade the RNA 5′ to 3′ toward the RNA polymerase. When the exonuclease catches up with the polymerase, transcription stops. Subsequently, polyadenylate polymerase enzymatically adds 20 to 200 adenine nucleotides to the 3′ end of the new transcripts. Polyadenylation is important for the localization, stability, and translation of mRNA in the cell. Almost all pol II transcripts are polyadenylated.

Advanced Concepts

Some pol II transcripts have alternative polyadenylation sites. Termination of the transcriptions at alternative locations on the transcript produces different transcripts (isoforms) from the same gene.[6]

Pol I terminates transcription just prior to a site in the DNA (Sal box) with the cooperation of a termination factor, TTF1.[6] The pol III termination signal is a run of adenine residues in the template. Pol III transcription termination requires a termination factor.

REGULATION OF TRANSCRIPTION

Gene expression is a key determinant of phenotype. The sequences and factors controlling when and how much protein is synthesized are equally as important as the DNA sequences encoding the amino acid makeup of a protein. Early studies aimed at the characterization of gene structure were confounded by phenotypes that resulted from aberrations in gene expression rather than in protein structural alterations.

Advanced Concepts

With advances in crystallography, the molecular mechanisms of RNA synthesis were revealed. During transcription, DNA enters a positively charged cleft between the two largest subunits of the RNA polymerase. At the floor of the cleft is the active site of the enzyme to which nucleotides are funneled through a pore in the cleft beneath the active site (pore 1). In the active site, the DNA strands are separated and the RNA chain is elongated, driven by cleavage of phosphates from each incoming ribonucleoside triphosphate. The resulting DNA–RNA hybrid moves out of the active site nearly perpendicular to the DNA coming into the cleft. After reaching a length of 10 bases, the newly synthesized RNA dissociates from the hybrid and leaves the complex through an exit channel. Three protein loops—“rudder,” “lid,” and “zipper”—are involved in hybrid dissolution and exit of the RNA product.[7,8]

For some genes whose products are in continual use by the cell, gene expression is constant or **constitutive.** For other genes, expression is tightly regulated throughout the life of the cell. Because gene products often function together to bring about a specific cellular response, specific combinations of proteins in stoichiometric balance are crucial for cell differentiation and development. Protein availability and function are controlled at the levels of transcription, translation, and protein modification and stability. The most immediate and well-studied level of control of gene expression is transcription initiation. Molecular technology has led to an extensive study of transcription initiation, so a large amount of information on gene expression refers to this level of transcription.

Regulation of Messenger RNA Synthesis at Initiation

Two types of factors are responsible for regulation of mRNA synthesis: cis factors and trans factors (transcription factors; Fig. 2.2). Cis factors are DNA sequences that mark places on the DNA involved in the initiation and control of RNA synthesis. Trans factors are proteins that bind to the cis sequences and direct the assembly of transcription complexes at the proper gene. In order for transcription to occur, several proteins must assemble at the gene’s transcription initiation site, including specific and general transcription factors and the RNA polymerase complex.

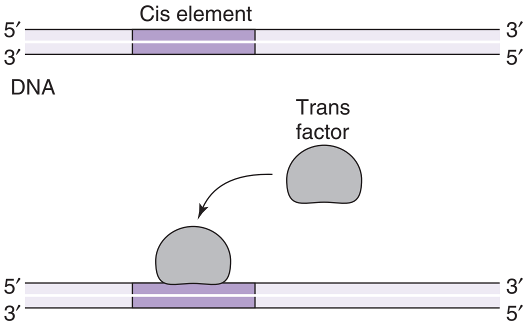


FIGURE 2.2 Cis elements are DNA sequences that are recognized by regulatory proteins (trans factors). Binding of trans factors can turn gene expression on or off.

An operon is a series of structural genes transcribed together on one mRNA and subsequently separated into individual proteins. In organisms with small genomes, such as bacteria and viruses, operons bring about the coordinated expression of proteins required at the same time, for example, the enzymes of a metabolic pathway. For example, the lactose operon (lac operon) in *Escherichia coli* contains three structural genes: *lacZ, lacY,* and *lacA,* which are all required for the metabolism of lactose. The *lacZ* gene product, β-galactosidase, hydrolyzes lactose into glucose and galactose. The *lacY* gene product, lactose permease, transports lactose into the cell. The *lacA* gene product, thiogalactoside transacetylase, transacetylates galactosides. A *lacI* gene encodes a protein repressor that binds to the *lacO* cis factor in the DNA (the operator) just 5′ to the start of the operon near where RNA polymerase binds *(lacP).* When *E. coli* is growing on glucose as a carbon source, the lactose-metabolizing enzymes are not required, and this operon is minimally expressed. Within two to three minutes after shifting to a lactose-containing medium, the expression of these enzymes is increased a thousand-fold.

Historical Highlights

The production of particular proteins by bacteria growing in media containing specific substrates was observed early in the last century, a phenomenon termed enzyme adaptation, later called induction. Detailed analysis of the lactose operon in *E. coli* was the first description of an inducible gene expression at the molecular level. The effect of gene expression on phenotype was initially demonstrated by Monod and Cozen-Bizare in 1953 when they showed that synthesis of tryptophan in *Aerobacter* was inhibited by tryptophan.[9] Jacob and Monod subsequently introduced the concept of two types of genes, structural and regulatory, in the lactose operon in *E. coli*.[10]

Figure 2.3 shows a map of the lac operon. The three structural genes of the operon are copied into a single transcript under the control of the operator.

The sequences coding for the repressor protein are located just 5′ to the operon. In the absence of lactose, the repressor protein binds to the operator sequence and prevents transcription of the operon (Fig. 2.4A). When lactose is present, it binds to the repressor protein and changes its conformation and lowers its affinity to bind the operator sequence. This results in expression of the operon (Fig. 2.4B). Jacob and Monod originally deduced these details through analysis of a series of mutants (changes in the DNA coding for the various components of the operon).[11] Since their work, numerous regulatory systems have been described in prokaryotes and eukaryotes, all using the same basic idea of combinations of cis and trans factors.

Other operons are controlled in a similar manner by the binding of regulatory trans factors to cis sequences preceding the structural genes (Fig. 2.5A). A different

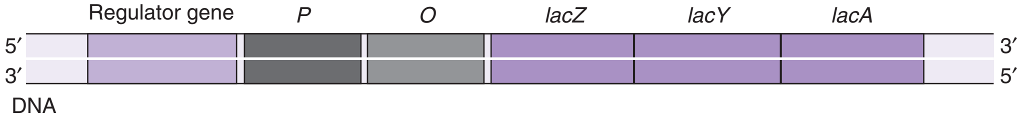


FIGURE 2.3 General structure of the lac operon. The regulator or repressor gene codes for the repressor protein trans factor that binds to the operator.

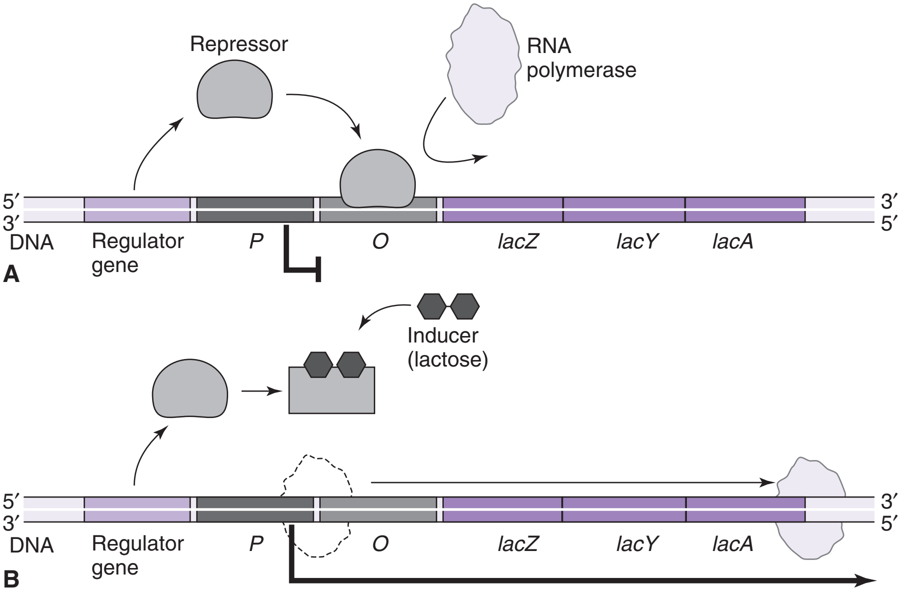

Repressor
RNA polymerase
5′
3′
DNA
Regulator gene
P
O
lacZ
lacY
lacA
3′
5′
A
Inducer (lactose)
5′
3′
DNA
Regulator gene
P
O
lacZ
lacY
lacA
3′
5′
B

FIGURE 2.4 Two states of the lac operon. (A) The repressor protein *(R)* binds to the operator cis element *(O)* preventing transcription of the operon from the promoter *(P)*. (B) In the presence of the inducer lactose, the inducer binds to the repressor, changing its conformation and decreasing its affinity for the operator, allowing transcription to occur.

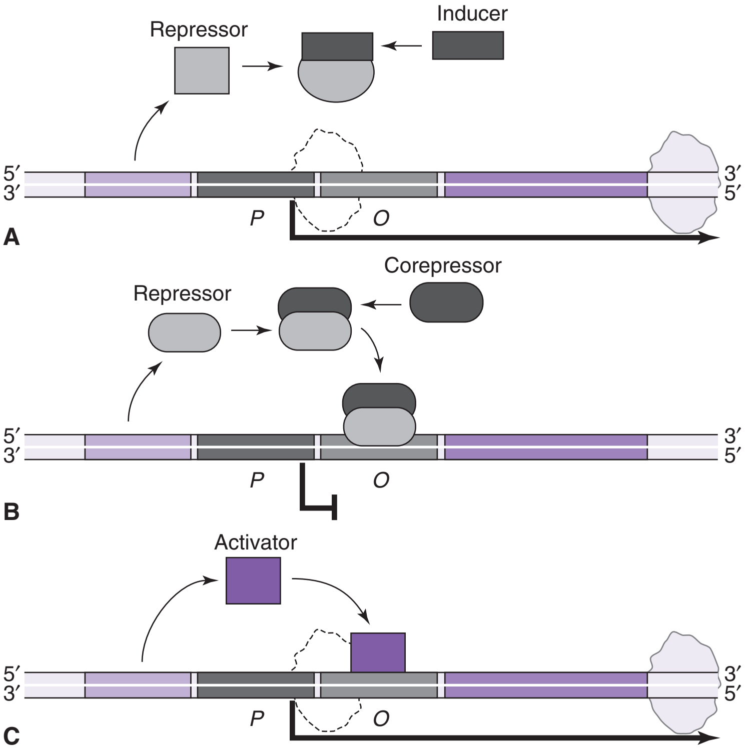

Repressor
Inducer
5′
3′
P
O
3′
5′
A
Corepressor
Repressor
5′
3′
P
O
3′
5′
B
Activator
5′
3′
P
O
3′
5′
C

FIGURE 2.5 Modes of regulation in prokaryotes include induction as found in the lac operon (A), repression as found in the arg operon (B), and activation as in the mal operon (C).

type of negative control is that found in the arg operon, where a corepressor must bind to a repressor in order to turn off transcription (**enzyme repression**; Fig. 2.5B). Compare this with the inducer that prevents the repressor from binding the operator to turn on expression of the lac operon **(enzyme induction).** The mal operon is an example of positive control where an activator binds with RNA polymerase to turn on transcription (Fig. 2.5C).

Another mechanism of control in bacteria is attenuation. This type of regulation works through the formation of stems and loops in the RNA transcript by intrastrand hydrogen bonding of complementary bases (Fig. 2.6). These secondary structures allow or prevent transcription, for instance, by respectively exposing or sequestering ribosome-binding sites at the beginning of the transcript.

The general arrangement of cis factors in prokaryotes and eukaryotes is shown in Figure 2.7. These sequences are usually 4 to 20 base pairs (bp) in length. Some are inverted repeats with the capacity to form a cruciform structure in the DNA duplex recognizable by specific proteins. Prokaryotic regulatory sequences are usually found within close proximity of the gene. Eukaryotic genes have both proximal and distal regulatory elements. Distal eukaryotic elements can be located thousands of base pairs away from the genes they control. Enhancers and silencers are examples of distal regulatory elements that, respectively, stimulate or dampen the expression of distant genes (Fig. 2.8).

Advanced Concepts

Unlike prokaryotes, eukaryotes do not have operons. Coordinately expressed genes can be scattered in several locations. Synchronous expression is brought about in eukaryotes by combinatorial control; that is, genes that are expressed in a similar pattern share similar cis elements so that they respond simultaneously to specific combinations of controlling trans factors.

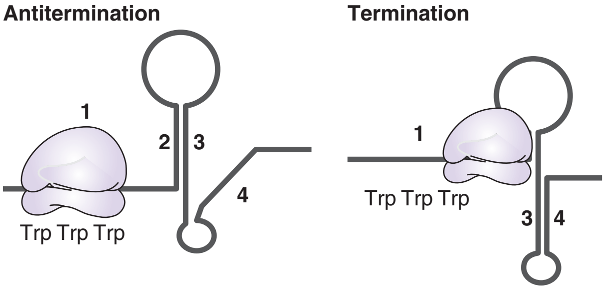


FIGURE 2.6 Attenuation regulates expression at the level of translation. In addition to repression of RNA synthesis, enzymes encoded in the trp operon are also regulated at the level of translation. With low tryptophan, the ribosome pauses at trp codons in the mRNA, allowing the anti-terminator hairpin to form by hybridization of RNA regions 2:3, and translation continues. When tryptophan levels are adequate, the ribosome moves quickly through the trp codons, inducing hybridization of RNA regions 3:4 and termination of translation.

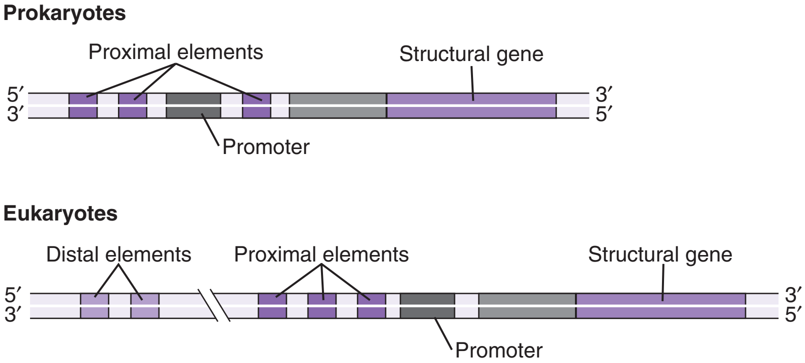


FIGURE 2.7 Cis regulatory elements in prokaryotes are located close to the structural genes they control in the vicinity of the promoter. In eukaryotes, distal elements can be located thousands of base pairs away from the genes they control. Proximal elements can be located in or around the genes they control. Elements may also be located behind their target genes.

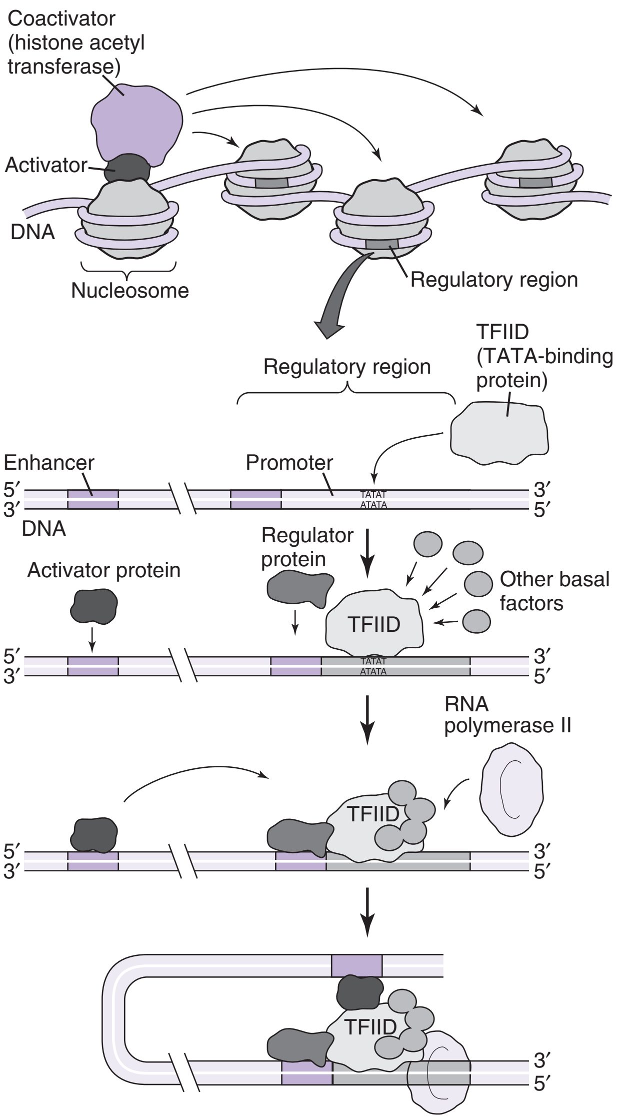

Coactivator
(histone acetyl
transferase)
Activator
DNA
Nucleosome
Regulatory region
Regulatory region
TFIID
(TATA-binding
protein)
Enhancer
Promoter
5′
3′
TATAT
ATATA
3′
5′
DNA
Activator protein
Regulator
protein
Other basal
factors
TFIID
5′
3′
TATAT
ATATA
3′
5′
RNA
polymerase II
TFIID
5′
3′
3′
5′
TFIID
3′
5′

FIGURE 2.8 Interaction of transcription factors and histone acetylation at the regulatory region of a gene *(top)* induce assembly of protein factors and RNA polymerase at the promoter *(bottom).*

Post-Transcriptional Regulation

For most genes, the RNA transcript has to be translated to protein to bring about phenotype. The transcript must be processed and translated into protein. Several factors affect the stability of the RNA transcript, including RNA sequence structure and the presence of exonucleases and endonucleases that digest the RNA. Enzymatic and structural alterations in RNA interfere with its processing in eukaryotes and its translation into protein.[12]

In eukaryotes, RNA processing steps will affect the production of protein and ultimately the phenotype. Alternate splicing in combination with protein factors is responsible for many tissue- and development-specific gene-expression patterns (e.g., during hematopoiesis).[13]

Advanced Concepts

Circular RNA structures (circular RNA) are implicated in the control of gene expression, among a variety of other cellular functions.[14,15] These RNAs, first thought to be viral products, are endogenous by-products of RNA splicing in protein-coding genes. Their roles include protein binding, activation of transcription, and subcellular localization. They are naturally stable because they do not have ends subject to exonucleases. They are implicated in tissue development, stress response, cancer, and other diseases.[13,16]

In prokaryotes, RNA transcription and translation are concurrent. This protects the RNA from the processing and exogenous factors. Just as in eukaryotes, however, RNA stability in prokaryotes is affected by secondary structure (folding of the RNA molecule) and polyadenylation of the 3′ end of the transcripts.[17] Codon usage and cofactor availability may also alter the speed of translation, resulting in decreased or increased amounts of protein.[18]

Post-Translational Regulation

Once proteins are synthesized, the genes might be considered successfully expressed; however, post-translational events influence the function of the completed protein. Some nascent peptides and proteins are modified with lipids, sugars, and other factors by cellular enzymes in order to function. Protein-modifying enzymes include phosphatases, kinases, ubiquitinases, transferases, acetylases, methylases, and ligases. As proteins move through the endoplasmic reticulum and Golgi, they are directed to their functional locations or marked for secretion. By controlling protein function, these activities are important for cell phenotype and behavior. Signals in Golgi have even been implicated in oncogenic transformation.[19]

Gene expression is measured in the laboratory at the level of transcription and protein production. Single or multiple genes are tested, depending on the nature of the test or study. The range of activities and interactions that participate in the production and function of RNA and proteins complicate the interpretation of gene expression and its control. In addition to expression levels, changes in the DNA sequence (mutations) can result in successfully expressed and modified transcripts or proteins without function. This and other possibilities must be taken into account when interpreting gene-expression results.

EPIGENETICS

In 1942, Conrad Waddington, an embryologist, defined **epigenetics** as a developmental phenomenon that allowed cells to take on different phenotypes (differentiate) without changes in the genetic structure. Epigenetics was a way to explain how a single cell could differentiate into a multicellular organism without vast genotypic changes. He related this idea to the notion of "epigenesis," the 17th-century concept of the complex interactions between Mendel's genotype and phenotype.[20] Sixteen years later, microbiologist David Nanney proposed an alternate two-part system of cellular control, one based on genetics (DNA) and the other determining which genes would be expressed at which time.[21] In this model the term *epigenetics* was used to emphasize the reliance of the auxiliary system on the genetic information. With the subsequent growth in understanding of the nature of the regulation of gene expression, epigenetics was further defined as mitotically and meiotically heritable changes in phenotype not encoded in genotype. Epigenetics may be a mechanism of rapid and heritable adaptation to environmental changes without alteration in the DNA genotype. The current study of epigenetics involves chemical and structural modifications in chromatin and the activity of particular noncoding RNAs. Changes in gene expression and phenotype as a result of aberrant epigenetic phenomena are studied in a variety of disease states. These phenomena include histone modification, nucleic acid methylation, and noncoding RNA.

Histone Modification

Chromatin is nuclear DNA and its associated proteins. In eukaryotes, the double helix of chromosomal DNA is compacted onto nucleosomes. A **nucleosome** is about 150 bases of DNA wrapped around a complex of eight histone proteins, two each of histone 2A, histone 2B, histone 3, and histone 4. Histones are not only structural proteins but also regulate access of trans factors and RNA polymerase to the DNA helix (Fig. 2.9). Modification of histone proteins affects the activity of chromatin-associated proteins and transcription factors that increase or decrease gene expression. Through these protein interactions, chromatin can move between transcriptionally active and transcriptionally silent states. Different types of histone modifications, including methylation, phosphorylation, ubiquitination, and acetylation, establish a complex system of control. Modifications have specific effects on chromatin; for example, acetylation lowers the positive charge of the histones, decreasing binding strength to the negatively charged DNA, thereby making the DNA more available for interaction with transcription factors and RNA polymerase (open chromatin). Methylation of histones attracts enzymes that further methylate DNA, resulting in decreased gene expression (closed chromatin). Phosphorylation, ubiquitination, and other modifications directly affect the histone–DNA interaction or recruit other modifying enzymes (Table 2.3).

Modifications occur on specific amino acids, and carboxy ends the histone protein sequences (Fig. 2.10). These "histone tails" extend from the nucleosome, where they are available to the acetylases, methylases, and other enzymes. The histone modifications (histone marks) form an environment for interaction with other chromatin-binding factors and, ultimately, regulatory proteins.[22] A "histone code" has been proposed, analogous to the genetic code, correlating specific histone modifications with biological effects (Table 2.1).[23,24] The recommended nomenclature for core histone (H2A, H2B, H3, H4) modifications is the name of the histone followed by the amino acid and its position in the protein using the single-letter code, followed by the modification. Acetylation of lysine at position 27 of histone H3 would be designated H3K27Ac. Methylation of lysine at position 20

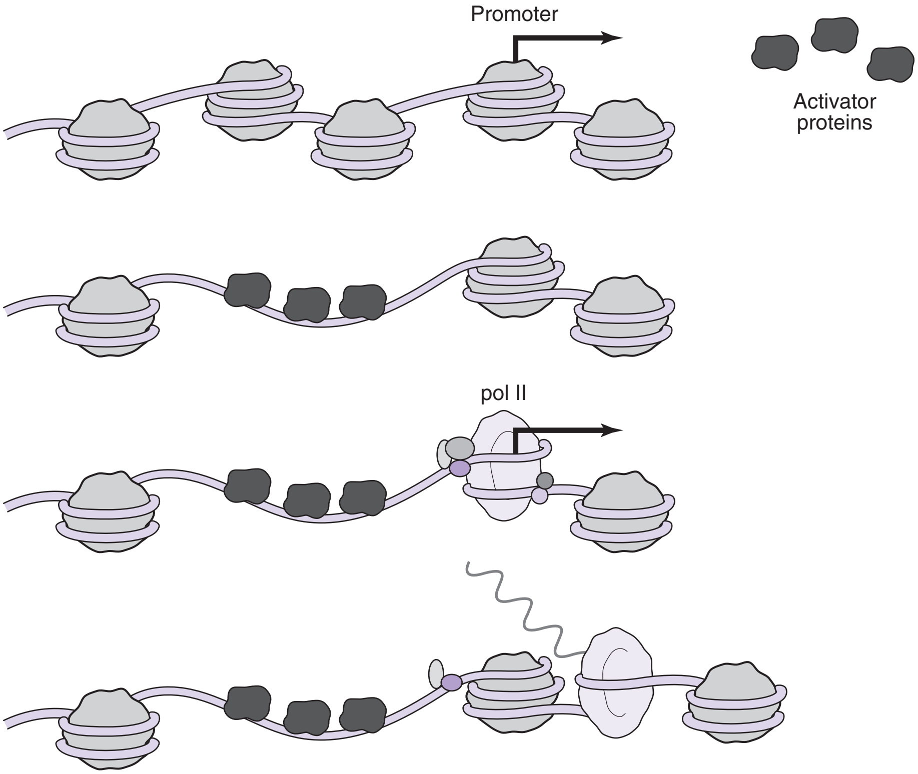


FIGURE 2.9 Histones are components of nucleosomes required for organization of DNA in the nucleus. They are also important regulators of gene expression. In order for trans factors, such as activators, RNA polymerase, and its cofactors, to access DNA, histones must release the cis sequences and promoter to which they bind.

of histone H2B would be H2BK20Me.The involvement of histone modification with gene expression has led to the study of aberrations in these modifications in disease states such as viral infections and neoplastic cells. Therapeutic agents such as histone deacetylase inhibitors are currently in use for some hematological malignancies.[25] Thus, histone modifications may be another target for diagnostic, prognostic, and therapeutic applications.

Nucleic Acid Methylation

DNA Methylation

DNA methylation is another type of epigenetic regulation of gene expression in eukaryotes and prokaryotes. In vertebrates, methylation of DNA occurs in cytosine-guanine–rich sequences in the DNA (CpG islands; Fig. 2.11). CpG islands were initially defined as regions of more than 200 bp in length with an observed/expected ratio of the occurrence of CpG of greater than 0.6.[26] This definition may be modified to a more selective GC content to exclude unrelated regions of naturally high GC content.[27] CpG islands are found around the first exons, promoter regions, and sometimes toward the 3′ ends of genes. There are over 45,000 CpG islands in the human genome. Aberrant DNA methylation at these sites is a source of dysregulation of genes in disease states. Methylation of cytosine residues in the promoter regions of tumor-suppressor genes is a mechanism of inactivation of these genes in cancer.[28]

Methylation of DNA is the main mechanism of genomic **imprinting,** the stage- and gamete-specific silencing of genes.[29] Imprinting maintains the balanced expression of genes in growth and embryonic development by selective methylation of **homologous** genes. This controlled methylation occurs during gametogenesis and

TABLE 2.3 The Histone Code

Histone	Modification	Function
H2A	Acetylation	Activation
H2A	Phosphorylation	Mitosis, DNA repair*
H2A	Ubiquitylation	Stem cells
H2B	Acetylation	Activation
H2B	Phosphorylation	Apoptosis
H2B	Ubiquitylation	Transcription
H3	Unmodified	Silencing
H3	Methylation	Silencing/activation*
H3	Phosphorylation	Mitosis/activation*
H3	Acetylation	Activation/histone positioning*
H4	Phosphorylation	Activation
H4	Methylation	Silencing/activation*
H4	Acetylation	DNA repair/histone positioning

*Function depends on which amino acid is modified or if a combination of modifications is present.

is different in male and female gametes. A convenient illustration of imprinting is the comparison of mules and hinnies. A mule (progeny of a female horse and male donkey) has a distinct phenotype from that of a hinny (progeny of a male horse and a female donkey). The difference is due, in part, to the distinct imprinting of genes inherited through egg formation (oogenesis) versus those inherited through sperm formation (spermatogenesis). Genetic diseases in humans, such as Angelman syndrome and Prader–Willi syndrome, are clinically distinct conditions that result from the same genetic defect on chromosome 15.[30] The phenotypic differences depend on whether the genetic lesion involves the maternally or paternally inherited chromosome. Imprinting may be partly responsible for the abnormal development and phenotypic characteristics of cloned animals because the process of cloning by nuclear transfer bypasses egg and sperm formation (gametogenesis).[31,32]

There are four enzymes that methylate DNA in mammals: DNA methyltransferase 1 (DNMT1), DNA methyltransferase 3a/3b (DNMT3), and DNA methyltransferase 3L (DNMT3L). A fifth enzyme, DNA methyltransferase 2, may be involved in modifying bases in tRNA or other targets. DNMT1 is a maintenance enzyme and mainly methylates hemimethylated DNA, that is, cytosines that are methylated on one strand of the double helix and not the other. This is the state of methylated DNA after replication where the newly synthesized strand is not methylated like the parent strand. DNMT3 is a de novo methylase and methylates unmethylated DNA, establishing newly methylated regions, as in imprinting. DNMT3L is likely nonfunctional as a methylase because it lacks a functional catalytic site. It may regulate DNMT3A and 3B activities by occupying DNMT3A and 3B protein or DNA-binding sites.

DNMT1 and DNMT3 recognize cytosine bases in DNA followed by guanine bases (CG or CpG). The symmetrical nature of the CG sequence reflects the symmetry of the enzymes. The complement of 5′-CG-3′ is 5′-CG-3′. The enzymes are recruited to potentially methylated areas by proteins that bind modified histones as well as proteins that bind hemimethylated DNA.

Advanced Concepts

One theory holds that a preexisting and gene-specific **histone code** extends the information potential of the genetic code.[24] Accordingly, euchromatin, which is transcriptionally active, has more acetylated histones and less methylated histones than transcriptionally silent heterochromatin made up of more condensed nucleosome fibers. Although methylated histones can activate transcription by recruiting histone acetylases, the establishment of localized areas of histone methylation can also prevent transcription by recruiting proteins for DNA methylation and heterochromatin formation. This is one form of gene or transcriptional silencing.[22] Transcriptional silencing is responsible for inactivation of the human X chromosome in female embryo development and **position effects,** the silencing of genes when placed in heterochromatic areas.

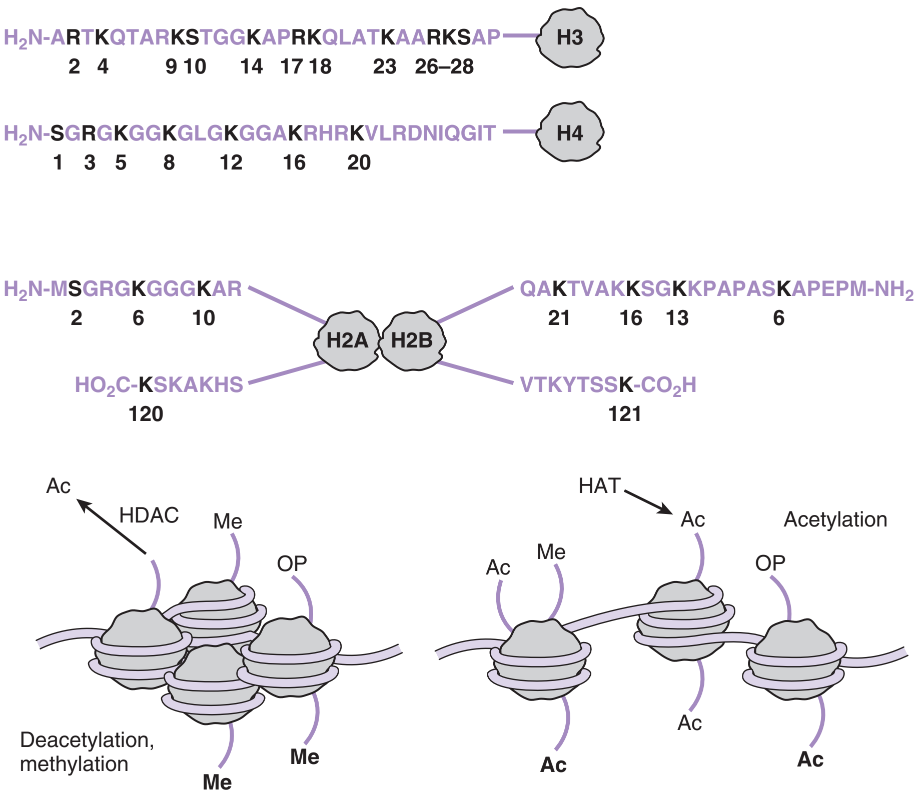

H2N-ARTKQTARKSTGGKAPRKQLATKAARKSAP
2 4 9 10 14 17 18 23 26–28
H3
H2N-SGRGKGGKGLGKGGAKRHRKVLRDNIQGIT
1 3 5 8 12 16 20
H4
H2N-MSGRGKGGGKAR
2 6 10
HO2C-KSKAKHS
120
H2A
H2B
QAKTVAKKSGKKPAPASKAPEPM-NH2
21 16 13 6
VTKYTSSK-CO2H
121
Ac
HDAC
Me
OP
Deacetylation, methylation
Me
Me
HAT
Ac
Acetylation
Ac
Me
OP
Ac
Ac
Ac

FIGURE 2.10 Histones H2A, H2B, H3, and H4 are modified at specific amino acids on the amino terminal and carboxyterminal ends *(top)*. Depending on the modifications (histone marks), the associated DNA will form closed, nonexpressed heterochromatin *(left)* or open, actively transcribed euchromatin *(right)*.

...GGAGGAGCGCGCGGCGGCGGCCAGAGA

AAAGCCGCAGCGGCGCGCGCGCACCCGGA

CAGCCGGCGGAGGCGGG...

FIGURE 2.11 CpG islands are sequences of DNA rich in the C-G dinucleotides. These structures have no specific sequence other than a higher-than-expected occurrence of CpG.

Advanced Concepts

Normally, DNA is unmethylated or transiently methylated in CpG islands before active genes. In the vast intergenic regions of DNA and in gene bodies (coding sequences), methylation is much heavier. Abnormal hypermethylation of CpG islands as well as hypomethylation of intergenic areas can occur. In some types of colon and brain cancer, hypermethylation of multiple CpG islands has been defined as a CpG Island Methylator Phenotype (CIMP) with diagnostic and prognostic implications.[33] Hypomethylation in the much larger intergenic areas results in an overall hypomethylation in tumor cells. Hypomethylation around retrotansposons such as the long interspersed nuclear elements (LINEs) is thought to be a cause of the genomic instability (chromosome alterations, breakage, and loss) seen in some cancers.

DNA Demethylation

In normal cells, DNA methylation, especially in CpG islands, is reversible. In this way, DNA methylation can serve as an additional layer of transcriptional control in addition to cis and trans genetic factors. DNA methylation occurs passively through rounds of DNA replication, enzymatic alteration, and repair or by active conversion.[34] The major pathway of active demethylation in vertebrate cells is a series of oxidative reactions catalyzed by the Ten Eleven Translocation (TET) enzymes. These enzymes first convert 5-methylcytosine to 5-hydroxymethylcytosine, which is then converted to 5-formylmethyl cytosine and then to 5-carboxymethyl cytosine. A decarboxylase enzyme catalyzes the final step to restore 5-methylcytosine.

Advanced Concepts

The process of demethylation catalyzed by the TET enzymes connects the epigenetic state of DNA with metabolic processes.[35] The TET enzymes require alpha-ketoglutarate in order to function. Alpha-ketoglutarate is a product of the isocitrate dehydrogenase enzymes (IDH1 and IDH2), which are part of the citric acid cycle. Specific amino acid substitution mutations in the IDH enzymes replace the normally produced a-kg with 2-hydroxy glucose. The 2-hg competes with a-kg, causing loss of function of the TET enzymes. Thus, components of the metabolic citric acid cycle can alter the methylation state of DNA. DNA alterations in the IDH enzymes are found in cancer (e.g., glioblastoma and acute myeloid leukemia).

RNA Methylation

There are more than 100 post-transcriptional modifications on RNA.[36] Transfer RNA has a number of altered and methylated bases, increasing the recognition of charging enzymes and codon–anticodon binding. Messenger RNA, ribosomal RNA, and noncoding RNA are also methylated. The methylated bases are most frequently N^6-methyladenosine (m^{6}A) in a three-base recognition site, GAC, and less commonly AAC. About 25% of mRNAs have m^{6}A clustered around stop codons, sometimes found in 5′ untranslated regions of the RNA but not in the 3′ polyA tails. Methyltransferase enzymes encoded by *METTL14* and *METTL3* genes recognize and methylate target RNA regions.[37] RNA methylation affects RNA stability, splicing, and translation.

CLASSIFICATION OF EPIGENETIC FACTORS

Histone and nucleic acid modifications that bring about epigenetic processes have been classified as writers and erasers. Methyltransferases, phosphorylases, acetylases, and other enzymes that modify histones, RNA, and DNA are the writers. Deacetylases, TET enzymes, and enzymes that reverse these modifications are classified as erasers. A third category of readers is the transcription factors, methyl-group–binding proteins and others that recognize the state of chromatin and act accordingly.

Another functional classification for epigenetics in cancer cells has also been proposed. This classification considers the interrelationship between genetic and epigenetic factors. Classes include modulators, modifiers, and mediators. Modulators are those activities that activate the epigenetic process. This category includes genetic factors such as *IDH1/2* and the *KRAS* and *TP53* cancer genes. Modifiers are the writers and erasers of the previous system that change or maintain the chromatin structure. Modulators are the readers, such as transcription factors, that ultimately bring about the effects of the chromatin state.

The latter system emphasizes the concept that epigenetics is not completely independent of genetic control factors and vice versa. The combination of genetic and epigenetic activities forms a complex system that controls gene expression and ultimately phenotype through development, growth, and survival responding to environmental influences.

NONCODING RNAs

Noncoding RNAs carry out their function as RNA and are transcribed but not translated into protein. Noncoding RNAs were thought to be “leaky” transcription from intergenic areas of the genome. These species of RNA are now known to have roles in gene regulation and even in information transfer across generations.[38] Noncoding RNAs are found in oocytes and spermatozoa.

Piwi RNA (piRNA), which affects transposon transcription in germ cells, is specific to gamete-mediated epigenetic inheritance.[39]

Noncoding RNAs are classified as small, intermediate, and long, depending on their length. Small noncoding RNAs, microRNAs (miRNAs), short interfering RNAs (siRNAs), and long noncoding RNAs (lncRNAs) are regulators of gene expression.

MicroRNAs

MicroRNAs (miRNAs) are regulatory RNAs, 17 to 27 nucleotides (nt) in length, derived from endogenous RNA hairpin structures (RNA folded into double-stranded states through intrastrand hydrogen bonds). MicroRNAs were discovered in the worm *Caenorhabditis elegans*. Two 22-nt RNAs were shown to contribute to the temporal progression of cell fates by triggering down-regulation of target mRNAs.[40–42] These RNA species were also called small temporal RNAs (stRNAs). Functionally, miRNAs control gene expression by pairing with imperfectly complementary sequences in mRNAs to inhibit translation. Many miRNAs have been identified in eukaryotic cells and viruses.[43–45] Bacteria have genes that resemble miRNA precursors; however, the full miRNA system has not been demonstrated in bacteria. Over 5,600 miRNAs have been identified in humans.[46]

MicroRNAs perform diverse functions in eukaryotic cells, affecting gene expression, cell development, and defense. Because production of miRNAs is strictly regulated as to time or stage of cell development, finding them is a technical challenge. Many of these species are only present in virally infected cells or after the introduction of foreign nucleic acid by transformation. Novel approaches are being applied to the discovery of rare miRNAs expressed in specific cell types at specific times.

miRNA dysregulation is associated with a variety of diseases, especially cancer, where miRNAs promote the tumor-cell phenotype **(oncomirs).** Molecular mimics of miRNA and molecules targeted at miRNAs **(antimirs)** are being tested as therapeutic agents.[47]

Small Interfering RNAs

Small interfering RNAs (siRNAs) are the functional intermediates of RNA interference (RNAi), a defense in eukaryotic cells against viral invasion. In a process that is not yet completely understood, double-stranded RNA (dsRNA) species are believed to originate from transcription of inverted repeats or by the activity of cellular or viral RNA–directed RNA polymerases. Biochemical analysis of RNA interference has revealed that these 21- to 22-nucleotide dsRNAs are derived from successive cleavage of dsRNAs that are 500 or more nucleotide **base pairs** in length.[48] The ribonuclease III enzyme, dicer, is responsible for the generation of siRNA and another small RNA, microRNA (see following discussion), from dsRNA precursors.[49] siRNAs are not normally found in animal cells. They are frequently used in the laboratory to artificially turn off the expression of specific genes. Research using siRNAs is a standard approach to exploring the function and activity of genes of interest.

Other Small RNAs

Since the late 1990s, increasing varieties of small RNAs (sRNAs) have been described in prokaryotes and eukaryotes, including tiny noncoding RNAs (tncRNAs, 20 to 22 b), small modulatory RNAs (smRNAs, 21 to 23 b), small nucleolar RNAs (snoRNAs), tmRNAs, guide RNAs (gRNAs), and others. In addition to RNA synthesis and processing, these molecules influence numerous cellular processes, including plasmid replication, bacteriophage development, chromosome structure, and development. These small, untranslated RNA molecules have been termed sRNAs in bacteria and noncoding RNAs (ncRNAs) in eukaryotes.

Advanced Concepts

The first demonstration of directed RNA interference in the laboratory occurred in the experiments of Fire et al. with *Caenorhabditis elegans*.[50] These investigators injected dsRNA into worms and observed dramatic inhibition of the genes that generated the RNA. Since then, siRNAs have been introduced into plants and animals, including human cells growing in culture. Injection of long dsRNA can kill human cells, but gene silencing can be achieved by the introduction of siRNAs or plasmids coding for their dsRNA. Libraries of siRNAs or DNA plasmids encoding them have been made that are complementary to over 8,000 human genes.[51] These genetic tools have potential applications not only in identifying genes involved in disease but also in the treatment of some of these diseases, particularly cancers where overexpression or abnormal expression of specific genes is part of the tumor phenotype.

Regulation of Gene Expression by MicroRNAs

Thousands of microRNAs encoded among genes in higher eukaryotes regulate gene expression post-transcriptionally by binding to the 3′ end of mRNA and preventing its translation into protein (Fig. 2.12A). MicroRNAs are transcribed as larger precursors (pre-miRNAs), which are cleaved into hairpins by RNase III–like endonucleases, dicer and Drosha. The hairpins are further digested into short, single-stranded RNA imperfectly complementary to the 3′ end of the target gene.[52] Hybridization between the microRNA and the mRNA prevents translation of the target RNA and eventually results in degradation of the mRNA.

Regulation of Gene Expression by RNA Interference

Another phenomenon that affects transcription is RNA interference, or RNAi. First discovered in the worm *C. elegans* in 1993, RNAi is not present in humans; however, it is a useful tool in the research laboratory to selectively eliminate expression of specific genes in vitro. RNAi is mediated through siRNAs.[53] Like microRNA, the siRNAs and other proteins assemble into an enzyme complex, called the RNA-induced silencing complex (RISC; see Fig. 2.12B). The RISC uses the associated siRNA to bind and degrade mRNA with sequences exactly complementary to the siRNA. Translation of specific genes can thus be inhibited. Alternatively, siRNAs may bind to specific sequences in already-transcribed homologous RNA, targeting them for degradation into more siRNAs. Although the regulatory mechanisms are similar, regulation by siRNA requires the introduction of foreign RNA into the cell, whereas microRNAs are encoded in the cell's DNA.[54]

Advanced Concepts

RNA interference has been utilized as a method to specifically inactivate genes in the research laboratory. Preferable to gene deletions or "knock-out" methods that take months to inactivate a single gene, RNAi can specifically inactivate several genes in a few hours.

Initially, this method did not work well in mammalian cells, due to a cellular response elicited by the introduction of long dsRNAs that turns off multiple genes and promotes cell suicide. Adjustments to methods, such as using shorter dsRNAs, have been demonstrated to work in mammalian cell cultures, but the efficiency of silencing will vary among cell lines and different experimental conditions.[55] This type of specific gene control in vitro may lead to directed control of transcriptional states, which would be useful in the clinical laboratory setting as well as the research laboratory. Because of the high specificity of siRNAs, RNAi has also been proposed as a manner of gene and viral therapy.[56] Silencing has been targeted to growth-activating genes such as the vascular endothelial growth factor in tumor cells. Small interfering RNA silencing may also be aimed at the HIV and influenza viruses. As with other gene-therapy methods, the stability and delivery of the therapeutic siRNAs are major challenges.[57]

Long Noncoding RNAs

Long noncoding RNAs (lncRNAs) are defined by their length of 200 to 100,000 bases. lncRNAs are important regulators of chromatin structure, affecting gene expression and disease processes. Genomic analysis has revealed from 15,000 to as many as 58,648 lncRNAs encoded in the human genome.[58,59]

lncRNAs function in several ways (Fig. 2.13).[60] Through secondary structure (folding into double-stranded domains), lncRNAs can bind proteins that normally would bind to cis sequences in DNA, preventing their regulatory action. Multiple domains or hairpins in the lncRNAs can act as a scaffold to bring proteins together. Double- and single-stranded domains can bind

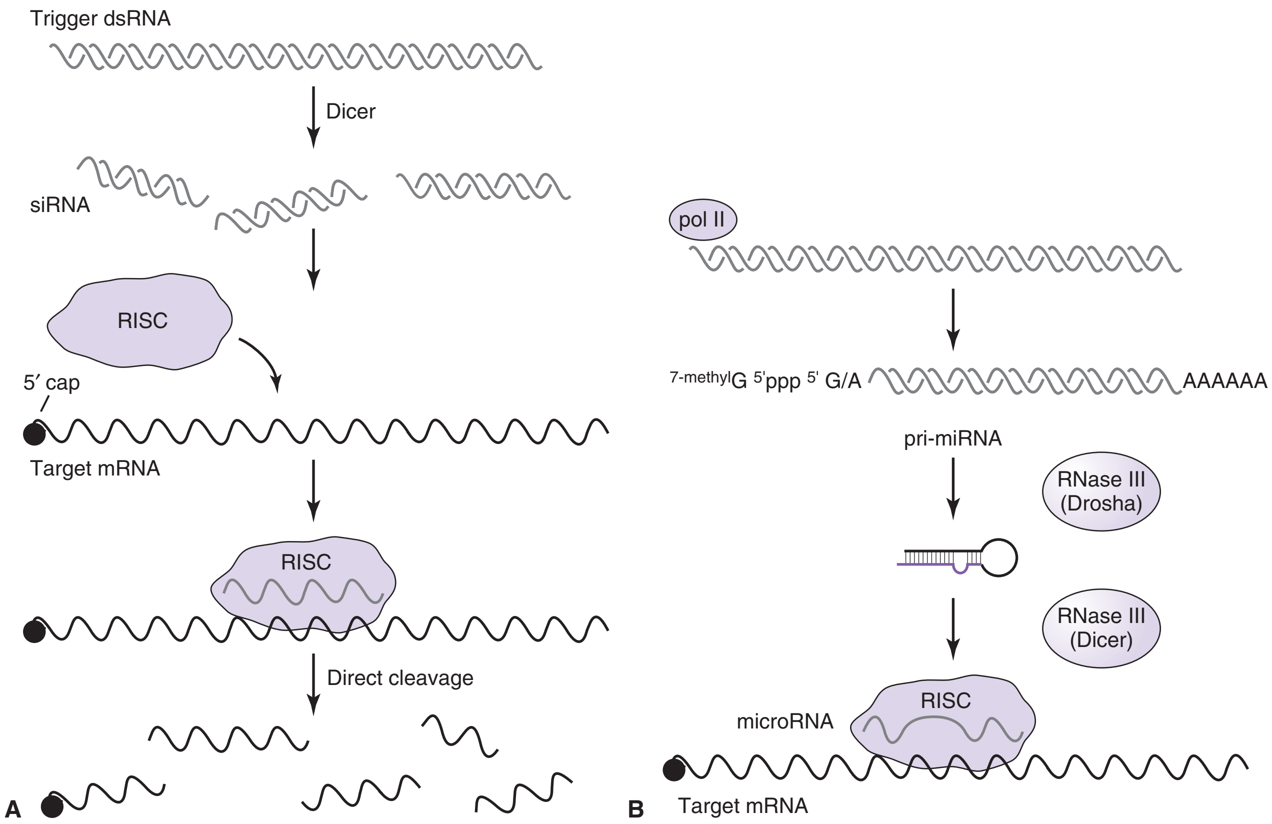


FIGURE 2.12 (A) In miRNA silencing, miRNA transcripts are transcribed from host genes, processed, and joined to the RNA-Induced Silencing Complex (RISC) assembly. (B) In siRNA silencing, a trigger double-stranded RNA is cleaved into siRNAs that become part of the RISC assembly. Led by the complementary siRNA sequences, RISC binds to the target RNA and begins RNA cleavage.

proteins to specific DNA-binding sites. Enhancers are cis sequences found far away from the genes that they control. Through scaffolding, lncRNAs may connect trans factors bound to the enhancer region to trans factors bound to proximal cis sites. lncRNAs may also bind small ligands, interact with miRNA in RISC complexes, form triplex structures with double-stranded DNA, or be further processed into small RNA fragments through RNAseP cleavage at the lncRNA 3′ ends.

ncRNAs are associated with disease states and are potential therapeutic targets.[61] The study of the biological roles of both small and long ncRNAs has led to efforts in the use of ncRNAs as therapeutic targets. The continued clarification of ncRNA mechanisms will lead to effective ncRNA drug design.

Historical Highlights

In 1904, Austrian zoologist Paul Kammerer wanted to show that animals acquire new characteristics when exposed to different environments and that these changes are heritable. He published many papers and books describing environmentally induced phenotypic changes passed down through multiple generations in amphibians and reptiles. These observations might have provided the first reports of epigenetic inheritance. Kammerer's work was controversial, however, with accusations of photographic image manipulation. Furthermore, there was limited success of other

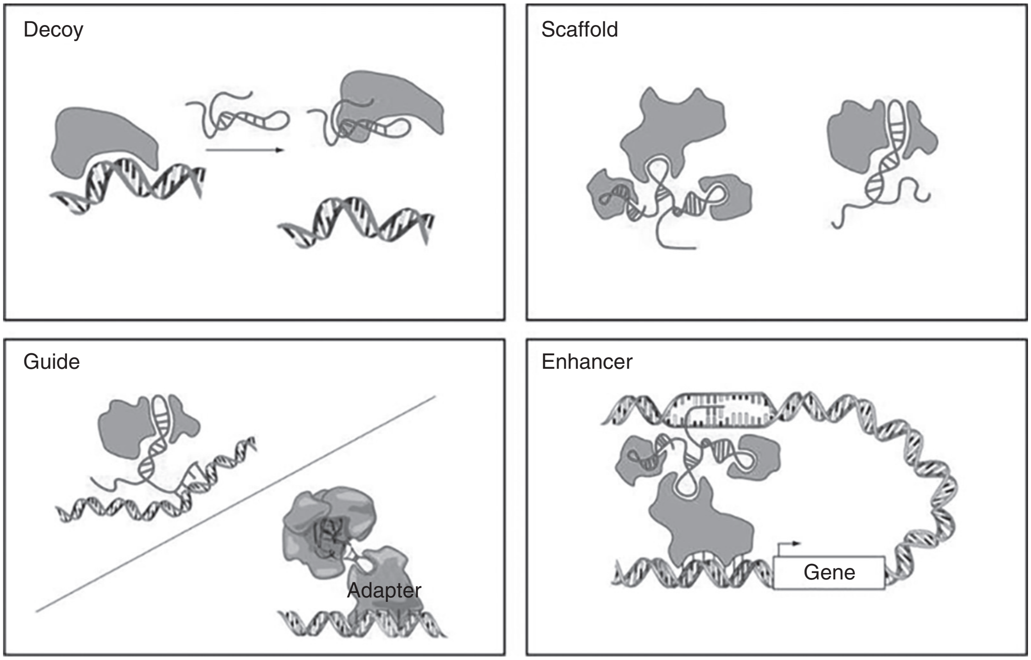


FIGURE 2.13 lncRNAs can titrate away DNA-binding proteins from cis sites or act as scaffolds to bring proteins into a complex. They may recruit transcription factors or histone-modification enzymes through RNA–DNA interactions or RNA interactions with DNA-binding proteins. lncRNAs may connect enhancer sites to proximal sites through protein–RNA and protein–DNA interactions. (Reprinted with permission, Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. *Annual Review of Biochemistry* 2012;81[1]:145–66.)

researchers in replicating Kammerer's findings on environmentally induced heritable foot-pad changes in toads and color changes in salamanders.[62] With the currently accepted molecular basis for epigenetics, researchers are now revisiting Kammerer's early ideas.[63]

Advanced Concepts

A special lncRNA, Xist (X-inactive specific transcript) is important for X-chromosome inactivation in mammalian female (XX) cells.[64] Along with other ncRNA and proteins, it is expressed from a region of the X chromosome, the X inactivation center (XIC). These products form a complex that initiates methylation of the chromosome. This mechanism silences nearly all genes on one X chromosome. Another lncRNA, Tsix, is complementary to Xist and extends across the Xist locus. Tsix negatively regulates Xist as an antisense RNA.[65]

STUDY QUESTIONS

Gene Expression

1. Proteins that bind to DNA to control gene expression are called __________ factors.

2. The binding sites on DNA for proteins that control gene expression are _________ factors.

3. How might a single mRNA produce more than one protein product?

4. The type of transcription producing RNA that is continually required and relatively abundant in the cell is called ______________________ transcription.

5. A set of structural genes transcribed together on one polycistronic mRNA is called a(n) _________.

The Lac Operon

Using the depiction of the lac operon in Figure 2.4, indicate whether gene expression (transcription) would be on or off under the following conditions: (P = promoter; O = operator; R = repressor).

- a. P+ O+ R+, no inducer present—OFF
- b. P+ O+ R+, inducer present—ON
- c. P− O+ R+, no inducer present—
- d. P− O+ R+, inducer present—
- e. P+ O− R+, no inducer present—
- f. P+ O− R+, inducer present—
- g. P+ O+ R−, no inducer present—
- h. P+ O+ R−, inducer present—
- i. P− O− R+, no inducer present—
- j. P− O− R+, inducer present—
- k. P− O+ R−, no inducer present—
- l. P− O+ R−, inducer present—
- m. P+ O− R−, no inducer present—
- n. P+ O− R−, inducer present—
- o. P− O− R−, no inducer present—
- p. P− O− R−, inducer present—

Epigenetics

1. Indicate whether the following events would increase or decrease the expression of a gene:
   - a. Methylation of cytosine bases 5′ to the gene
   - b. Histone acetylation close to the gene
   - c. siRNAs complementary to the gene transcript

2. How does the complementarity of siRNA to its target mRNA differ from that of miRNA?

3. What is imprinting in DNA?

4. What sequence structures in DNA, usually found 5′ to structural genes, are frequent sites of DNA methylation?

5. What is the RISC?

6. Name four functions of lncRNAs.

References

1. Werner M, Thuriaux P, Soutourina J. Structure-function analysis of RNA polymerases I and III. *Current Opinion in Structural Biology* 2009;19:740–745.
2. Machyna M, Heyn P, Neugebauer KM. Cajal bodies: where form meets function. *Wiley Interdisciplinary Reviews* 2013;4:17–34.
3. Jin D, Mata Martin C, Sun Z, Cagliero C, Zhou YN. Nucleolus-like compartmentalization of the transcription machinery in fast-growing bacterial cells. *Critical Reviews in Microbiology* 2017;52:96–106.
4. Tollervey D. Termination by torpedo. *Nature* 2004;432:456–457.
5. Epstein V, Dutta D, Wade J, Nudler E. An allosteric mechanism of Rho-dependent transcription termination. *Nature* 2010;463:245–249.
6. Kuhn A, Bartsch I, Grummt I. Specific interaction of the murine transcription termination factor TTF I with class-I RNA polymerases. *Nature* 1990;344:559–562.
7. Cramer P, Bushnell DA, Kornberg R. Structural basis of transcription: RNA polymerase II at 2.8 Angstrom resolution. *Science* 2001;292:1863–1876.
8. Gnatt A, Cramer P, Fu J, Bushnell DA, Kornberg R. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution. *Science* 2001;292:1876–1881.
9. Monod J, Cozen-Bizare G. L'effect d'inhibition specifique dans la biosynthese de la tryptophane-desmase chez *Aerobacter aerogenes*. *Comptes rendus de l'Académie des Sciences* 1953;236:530–532.
10. Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology* 1961;3:318–356.
11. Swerdlowa S, Schaapera RM. Mutagenesis in the lacI gene target of *E. coli:* improved analysis for lacId and lacO mutants. *Mutation Research* 2014;77:79–84.
12. Ignatova Z, Narberhaus F. Systematic probing of the bacterial RNA structurome to reveal new functions. *Current Opinion in Microbiology* 2017;36:14–19.
13. Conboy J. RNA splicing during terminal erythropoiesis. *Current Opinion in Hematology* 2017;24(3):215–221.
14. Fischer J, Leung AK. CircRNAs: a regulator of cellular stress. *Critical Reviews on Biochemistry and Molecular Biology* 2017;17:1–17.

15. Cortés-López M, Miura P. Emerging functions of circular RNAs. *Yale Journal of Biology and Medicine* 2016;89:527–537.

16. Ebbesen K, Hansen TB, Kjems J. Insights into circular RNA biology. *RNA Biology* 2016;16:1–11.

17. Rorbach J, Bobrowicz A, Pearce S, Minczuk M. Polyadenylation in bacteria and organelles. *Methods in Molecular Biology* 2014;1125:211–227.

18. Hajjari M, Sadeghi I, Salavaty A, Nasiri H, Birgani MT. Tissue specific expression levels of apoptosis involved genes have correlations with codon and amino acid usage. *Genomics and Informatics* 2016;14:234–240.

19. Makowski S, Tran TT, Field SJ. Emerging themes of regulation at the Golgi. *Current Opinion in Cell Biology* 2017;45:17–23.

20. Deichmann U. Epigenetics: the origins and evolution of a fashionable topic. *Developmental Biology* 2016;416:249–254.

21. Nanney D. Epigenetic control systems. *Proceedings of the National Academy of Sciences* 1958;44:712–717.

22. Bird A. Methylation talk between histones and DNA. *Science* 2001;294:2113–2115.

23. Berger S. The histone modification circus. *Science* 2001;292: 64–65.

24. Jenuwein T, Allis CD. Translating the histone code. *Science* 2001;293:1074–1079.

25. Fredly H, Gjertsen BT, Bruserud O. Histone deacetylase inhibition in the treatment of acute myeloid leukemia: the effects of valproic acid on leukemic cells, and the clinical and experimental evidence for combining valproic acid with other antileukemic agents. *Clinical Epigenetics* 2013;5:1–13.

26. Gardiner-Garden M, Frommer M. CpG islands in vertebrate genomes. *Journal of Molecular Biology* 1987;196:261–282.

27. Jones P, Takai D. The role of DNA methylation in mammalian epigenetics. *Science* 2001;293:1068–1070.

28. Tirado-Magallanes R, Rebbani K, Lim R, Pradhan S, Benoukraf T. Whole genome DNA methylation: beyond genes silencing. *Oncotarget* 2017;8:5629–5637.

29. Soellner L, Begemann M, Mackay DJ, Grønskov K, Tümer Z, Maher ER, Temple IK, Monk D, Riccio A, Linglart A, Netchine I, Eggermann T. Recent advances in imprinting disorders. *Clinical Genetics* 2017;91:3–13.

30. Cheon C. Genetics of Prader–Willi syndrome and Prader–Will-like syndrome. *Annals of Pediatric Endocrinology & Metabolism* 2016;21:126–135.

31. Solter D. Lambing by nuclear transfer. *Nature* 1996;380:24–25.

32. Yang L, Chavatte-Palmer P, Kubota C, O'Neill M, Hoagland T, Renard JP, Taneja M, Yang X, Tian XC. Expression of imprinted genes is aberrant in deceased newborn cloned calves and relatively normal in surviving adult clones. *Molecular Reproduction and Development* 2005;71(4):431–438.

33. Miller B, Sánchez-Vega F, Elnitski L. The emergence of pan-cancer CIMP and its elusive interpretation. *Biomolecules* 2016;6: 1–14.

34. Bochtler M, Kolano A, Xu GL. DNA demethylation pathways: Additional players and regulators. *Bioassays* 2017;39(1):1–13.

35. Waterfall J, Killian JK, Meltzer PS. The role of mutation of metabolism-related genes in genomic hypermethylation. *Biochemical Biophysical Research Communications* 2014;455:16–23.

36. Schwartz S. Cracking the epitranscriptome. *RNA* 2016;22: 169–174.

37. Wang X, Huang J, Zou T, Yin P. Human m6A writers: two subunits, two roles. *RNA Biology* 2017;14(3):300–304.

38. Wang Y, Liu H, Sun Z. Lamarck rises from his grave: parental environment-induced epigenetic inheritance in model organisms and humans. *Biological Reviews of the Cambridge Philosophical Society* 2017;92(4):2084–2111.

39. Chen Q, Yan M, Cao Z, Li X, Zhang Y, Shi J, Feng GH, Peng H, Zhang X, Zhang Y, Qian J, Duan E, Zhai Q, Zhou Q. Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. *Science* 2016;351:397–400.

40. Lee R, Feinbaum RL, Ambros V. The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 1993;75:843–854.

41. Reinhart B, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G. The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* 2000;403:901–906.

42. Wightman B, Ha I, Ruvkun G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. *Cell* 1993;75:855–862.

43. Lee RC, Ambros, V. An extensive class of small RNAs i *Caenorhabditis elegans*. *Science* 2001;294:862–864.

44. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. Identification of novel genes coding for small expressed RNAs. *Science* 2001;294:853–858.

45. Pfeffer S, Zavolan M, Grasser FA, Chien M, Russo JJ, Ju J, John B, Enright AJ, Marks D, Sander C, Tuschl T. Identification of virus-encoded microRNAs. *Science* 2004;304:734–736.

46. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research* 2014;42:D68–D73.

47. Rupaimoole R, Slack FJ. MicroRNA therapeutics: towards a new era for the management of cancer and other diseases. *Nature Review Drug Discovery* 2016;16:203–222.

48. Hamilton A, Baulcombe D. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* 1999;286: 950–952.

49. Knight S, Bass BL. A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in *Caenorhabditis elegans*. *Science* 2001;293:2269–2271.

50. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 1998;391:806–811.

51. Novina C, Sharp PA. The RNAi revolution. *Nature* 2004;430: 161–164.

52. Daugaard I, Hansen TB. Biogenesis and function of ago-associated RNAs. *Trends in Genetics* 2017;33:208–219.

53. Matzke M, Matzke AJM, Kooter JM. RNA: guiding gene silencing. *Science* 2001;293:1080–1083.

54. Tuschl T, Weber K. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* 2001;411: 494–498.

55. Munkácsy G, Sztupinszki Z, Herman P, Bán B, Pénzváltó Z, Szarvas N, Győrffy B. Validation of RNAi silencing efficiency using gene array data shows 18.5% failure rate across 429 independent experiments. *Molecular Therapy Nucleic Acids* 2016;5:1–7.

56. Wall N, Shi Y. Small RNA: can RNA interference be exploited for therapy? *Lancet* 2003;362:1401–1403.

57. Zhong X, Zhang D, Xiong M, Zhang L. Noncoding RNA for cancer gene therapy. *Recent Results in Cancer Research* 2016;209:51–60.

58. Iyer M, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, Barrette TR, Prensner JR, Evans JR, Zhao S, Poliakov A, Cao X, Dhanasekaran SM, Wu YM, Robinson DR, Beer DG, Feng FY, Iyer HK, Chinnaiyan AM. The landscape of long noncoding RNAs in the human transcriptome. *Nature Genetics* 2015;47:199–208.

59. Li C, Chen Y. Targeting long non-coding RNAs in cancers: progress and prospects. *International Journal of Biochemistry and Cell Biology* 2013;45:1895–1910.

60. Rinn J, Chang HY. Genome regulation by long noncoding RNAs. *Annual Review of Biochemistry* 2012;81:145–166.

61. Adams B, Parsons C, Walker L, Zhang WC, Slack FJ. Targeting noncoding RNAs in disease. *Journal of Clinical Investigation* 2017;127:761–771.

62. Vargas A. Did Paul Kammerer discover epigenetic inheritance? A modern look at the controversial midwife toad experiments. *Journal of Experimental Zoology* 2009;312B.

63. Vargas A, Krabichler Q, Guerrero-Bosagna C. An epigenetic perspective on the midwife toad experiments of Paul Kammerer (1880–1926). *Journal of Experimental Zoology* 2016;328:179–192.

64. Goodrich L, Panning B, Leung KN. Activators and repressors: a balancing act for X-inactivation. *Seminars in Cell & Developmental Biology* 2016;56:3–8.

65. Stavropoulos N, Lu N, Lee JT. A functional role for Tsix transcription in blocking Xist RNA accumulation but not in X-chromosome choice. *Proceedings of the National Academy of Sciences* 2001;98:10232–102327.

Section II

Common Techniques in Molecular Biology

Chapter 3

Nucleic Acid Extraction Methods

Outline

- ISOLATION OF DNA
  - *Preparing the Sample*
    - Bacteria and Fungi
    - Viruses
    - Nucleated Cells in Suspension (Blood and Bone Marrow Aspirates)
    - Plasma
    - Tissue Samples
  - *DNA Isolation Chemistries*
    - Organic Isolation Methods
    - Inorganic Isolation Methods
    - Solid-Phase Isolation
  - *Proteolytic Lysis of Fixed Material*
    - Rapid Extraction Methods
    - Isolation of Mitochondrial DNA
- ISOLATION OF RNA
  - *Total RNA*
  - *Specimen Collection*
  - *RNA Isolation Chemistries*
    - Organic Isolation
    - Solid-Phase Isolation
    - Proteolytic Lysis of Fixed Material
    - Isolation of polyA (Messenger) RNA
- MEASUREMENT OF NUCLEIC ACID QUALITY AND QUANTITY
  - *Electrophoresis*
  - *Spectrophotometry*
  - *Fluorometry*
  - *Microfluidics*

Objectives

3.1 Compare and contrast organic, inorganic, and solid-phase approaches for isolating cellular and mitochondrial DNA.

3.2 Describe DNA isolation from minimal and challenging samples.

3.3 Compare and contrast organic and solid-phase approaches for isolating total RNA.

3.4 Distinguish between the isolation of total RNA and that of messenger RNA.

3.5 Describe the gel-based, spectrophotometric, and fluorometric methods used to determine the quantity and quality of DNA and RNA preparations.

3.6 Calculate the concentration and yield of DNA and RNA from a given nucleic acid preparation.

The purpose of nucleic acid extraction is to release the nucleic acid from the cell for use in subsequent procedures. Ideally, the target nucleic acid should be free of contamination with protein, carbohydrate, lipids, or other nucleic acids, that is, DNA free of RNA or RNA free of DNA. The initial release of the cellular material is achieved by breaking the cell wall (if present) and cell and nuclear membranes (cell lysis). Lysis must take place in conditions that will not damage the nucleic acid. Following lysis, the target material is purified, and then the concentration and purity of the sample can be determined.

ISOLATION OF DNA

Although Miescher first isolated DNA from human cells in 1869 by precipitation,[1] the early routine laboratory procedures for DNA isolation were developed from density-gradient centrifugation strategies. Meselson and Stahl used such a method in 1958 to demonstrate semiconservative replication of DNA.[2] Later procedures took advantage of solubility differences among chromosomal DNA, plasmids, and proteins in alkaline buffers. Large (greater than 50 kbp) chromosomal DNA and proteins cannot renature properly when neutralized in acetate at low pH after alkaline treatment, forming large aggregates instead. As a result, they precipitate out of solution. The relatively small plasmids return to their supercoiled state and stay in solution. Alkaline lysis procedures were used extensively for extraction of 1- to 50-kb plasmid DNA from bacteria during the early days of recombinant DNA technology.

Preparing the Sample

Nucleic acid is routinely isolated from human, fungal, bacterial, and viral sources in the clinical laboratory (Table 3.1). The initial steps in nucleic acid isolation depend on the nature of the starting material and the test method. Some test systems include sample-collection vessels and reagents that begin the nucleic isolation in transport to the laboratory. Details of sample collection are important because of the diversity of sample types and nucleic acid sources (viruses, bacteria, nucleated cells). Use of an improper collection tube or method will compromise the test results. Test validation should include the sample-collection methods.

TABLE 3.1 Yield of DNA From Different Specimen Sources[29-35]

Specimen	Expected Yield*
Specimens Adequate for Analysis Without DNA Amplification	
Blood[†] (1 mL, 3.5–10 × 10^6 WBCs/mL) Buffy coat[†] (1 mL whole blood)	50–200 µg
Bone marrow[†] (1 mL)	100–500 µg
Cultured cells (10^7 cells)	30–70 µg
Solid tissue[‡] (1 mg)	1–10 µg
Lavage fluids (10 mL)	2–250 µg
Mitochondria (10-mg tissue, 10^7 cells)	1–10 µg
Plasmid DNA, bacterial culture (100-mL overnight culture)	350 µg–1 mg
Bacterial culture (0.5 mL, 0.7 absorbance units)	10–35 µg
Feces[§] (1 mg; bacteria, fungi)	2–228 µg
Specimens Adequate for Analysis With DNA Amplification	
Serum, plasma, cerebrospinal fluid[‖] (0.5 mL)	0.3–3 µg
Dried blood (0.5- to 1-cm-diameter spot)	0.04–0.7 µg
Saliva (1 mL)	5–15 µg
Buccal cells (1 mg)	1–10 µg
Bone, teeth (500 mg)	30–50 µg
Hair follicles[#]	0.1–0.2 µg
Fixed tissue** (5–10 × 10-micron sections; 10 mm^2)	6–50 µ
Feces[††] (animal cells, 1 mg)	2–100 pg

*Yields are based on optimal conditions. Assays will vary in yield and purity of sample DNA.

[†]DNA yield will vary with WBC count.

[‡]DNA yield will depend on type and condition of tissue.

[§]Different bacterial types and fungi will yield more or less DNA.

[‖]DNA yield will depend on degree of cellularity.

[¶]Dried blood yield from paper is less than from textiles.

[#]Mitochondrial DNA is attainable from hair shafts.

**Isolation of DNA from fixed tissue is affected by the type of fixative used and the age and the preliminary handling of the original specimen.

[††]Cells in fecal specimens are subjected to digestion and degradation.

Advanced Concepts

In surveying the literature, especially early references, the starting material determined which DNA extraction procedure was used. Extraction procedures were also modified to optimize the yield of specific products. A procedure designed to yield plasmid DNA did not efficiently isolate chromosomal DNA and vice versa.

Bacteria and Fungi

Many of the early recombinant DNA experiments were performed with gram-negative bacteria. Methods used at that time are the basis for a variety of DNA isolation systems currently used. Toxic substances used in the older methods have been replaced with safer materials. Cell walls are not thick and can be lysed by high pH and detergents.

Some bacteria and fungi have tough cell walls that must be broken to allow the release of nucleic acid. Several enzyme products that digest cell wall polymers are commercially available. Alternatively, cell walls can be broken mechanically by grinding or by vigorously mixing with glass beads. Gentler enzymatic methods are less likely to damage chromosomal DNA and thus are preferred for methods involving larger chromosomal targets as opposed to plasmid DNA. Treatment with detergent (1% sodium dodecyl sulfate) and strong base (0.2 M NaOH) in the presence of Tris base, ethylenediaminetetraacetic acid (EDTA), and glucose can also break bacterial cell walls.

Boiling in dilute sucrose, Triton X-100 detergent, Tris buffer, and EDTA after lysozyme treatment releases DNA that can be immediately precipitated with alcohol (see following discussion). DNA extracted by boiling or alkaline (NaOH) procedures is denatured (single stranded) and may not be suitable for methods, such as restriction enzyme analysis, that require double-stranded DNA.

Commercial reagents designed for isolation of DNA in amplification procedures, such as polymerase chain reaction (PCR), are used for yeast, filamentous fungi, and gram-positive bacteria. The advantage of these types of extraction is their speed and simplicity.

Viruses

Viral DNA is held within free viruses or integrated into the host genome along with host DNA. Some procedures use cell-free specimens, such as plasma, for viral detection. Others may require concentration of viroids by centrifugation or other methods.

Nucleated Cells in Suspension (Blood and Bone Marrow Aspirates)

Nucleic acid in human blood or bone marrow comes mostly from white blood cells (WBCs). Anticoagulants added to the collected sample will prevent clotting, which will trap the WBCs. Serum from clotted blood is a rich source of proteins, lipids, and other molecules, but not nucleic acids. Free WBCs carrying nucleic acids and cell-free nucleic acids are available from plasma. WBCs in blood or bone marrow specimens are purified of red blood cells (RBCs) and other blood components by either differential density-gradient centrifugation or differential lysis.

For differential density-gradient centrifugation, whole blood or bone marrow mixed with isotonic saline is overlaid with Ficoll. Ficoll is a highly branched sucrose polymer that does not penetrate biological membranes. Upon centrifugation, the mononuclear WBCs (the desired cells for isolation of nucleic acid) settle into a layer in the Ficoll gradient that is below the less dense plasma components and above the polymorphonuclear cells and RBCs. The layer containing the mononuclear WBCs is removed from the tube and washed by rounds of resuspension and centrifugation in saline before proceeding with the nucleic acid isolation procedure.

Another method used to isolate nucleated cells takes advantage of the differential osmotic fragility of RBCs and WBCs. Incubation of whole blood or bone marrow in hypotonic buffer or water will result in the lysis of the RBCs before the WBCs. The WBCs are pelleted by centrifugation, leaving the empty RBC membranes (ghosts) and hemoglobin, respectively, in suspension and solution.

Plasma

Solid tumors and transplanted organs release cells, exosomes, and nucleic acid into the bloodstream.[3,4] Shed cells have the molecular characteristics of the shedding tissue or organ, as does cell-free nucleic acid. Exosomes are small vesicles (30 to 100 nm in diameter), which form by invagination and budding from the inside of cellular endosome vesicles and are secreted by living cells. They contain proteins, lipids, and nucleic acid.[5] They can be collected by centrifugation. With the high sensitivity achievable from amplification procedures, these sources of circulating nucleic acids can be detected for purposes of diagnostic and prognostic analyses. Such a test is called a liquid biopsy. Liquid biopsies may preclude surgical biopsies and allow serial biopsy testing. Isolation of cell-free nucleic acid requires procedures to concentrate the target nucleic acid before isolation. These procedures include solid-phase collection of nucleic acid on beads or columns (see following discussion). Circulating tumor cells can be collected on special columns and filters that will selectively bind to tumor cells, avoiding WBCs that may be present.

Tissue Samples

Fresh or frozen tissue samples are dissociated before DNA isolation procedures can be started. Grinding the frozen tissue in liquid nitrogen, homogenizing the tissue, or simply mincing the tissue using a scalpel disrupts the whole-tissue samples. Fixed, embedded tissue may be deparaffinized by soaking in xylene (a mixture of three isomers of dimethylbenzene). Less toxic xylene substitutes are also often used for this purpose. After xylene treatment, the tissue is rehydrated by soaking it in decreasing concentrations of ethanol. Alternatively, fixed tissue may be used directly without dewaxing.

Depending on the fixative, DNA in preserved tissue may be broken and/or cross-linked to varying extents.[6] Comparative studies have shown that buffered formalin is the least damaging among tested fixatives, with mercury-based fixatives, such as Bouin's and B-5, being the worst for DNA recovery (see Table 3.2).[7,8] DNA quality will also depend on how the tissue was handled prior to fixation and the length of time the tissue was in fixation. In general, DNA target products of 100 base pairs (bp) or less can consistently be obtained from fixed tissue. Extended digestion in proteinase K may yield longer fragments.

TABLE 3.2 Tissue Fixatives Influencing Nucleic Acid Quality[7]

Fixative	Relative Quality of Nucleic Acid	Average Fragment Fixative Size Range (kb)
10% buffered neutral formalin	Good	2.0–5.0
Acetone	Good	2.0–5.0
Zamboni's	Not as good	0.2–2.0
Clarke's	Not as good	0.8–1.0
Paraformaldehyde	Not as good	0.2–5.0
Methacarn	Not as good	0.7–1.5
Formalin–alcohol–acetic acid	Not as good	1.0–4.0
B-5	Less desirable	<0.1
Carnoy's	Less desirable	0.7–1.5
Zenker's	Less desirable	0.7–1.5
Bouin's	Less desirable	<0.1

DNA Isolation Chemistries

Organic Isolation Methods

After release of DNA from the cell, further purification requires removal of contaminating proteins, lipids, carbohydrates, and cell debris. For organic isolation, this is accomplished using a combination of high salt, low pH, and an organic mixture of phenol and chloroform. The combination readily dissolves hydrophobic contaminants, such as lipids and lipoproteins; collects cell debris and strips away most DNA-associated proteins (Fig. 3.1). Isolation of small amounts of DNA from challenging samples such as fungi can be facilitated by pretreatment with cetyltrimethylammonium bromide, a cationic detergent that efficiently separates DNA from polysaccharide contamination. To avoid RNA contamination, RNase, an enzyme that degrades RNA, can be added at this point. Alternatively, RNase may also be added to the resuspended DNA at the end of the procedure.

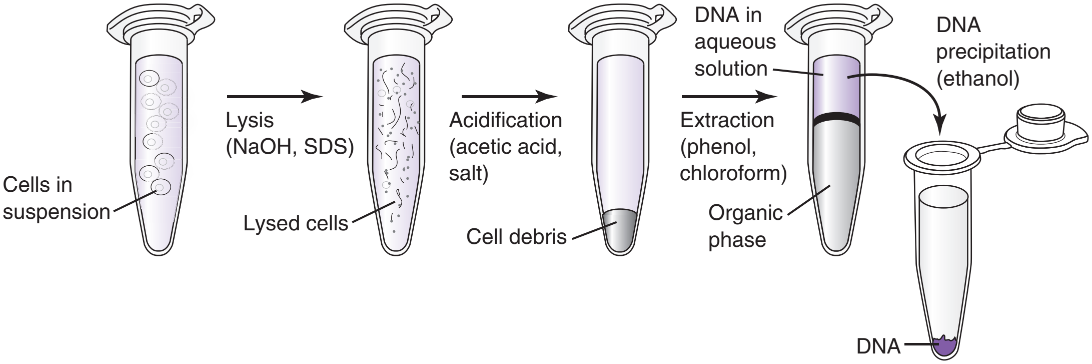


FIGURE 3.1 General scheme of organic DNA isolation. After lysis of cells, cellular contents are acidified, and particulate matter is pelleted and extracted with phenol and chloroform. This emulsion will separate into two phases. The upper aqueous phase contains the DNA that can be precipitated by mixing with ethanol. Collecting the precipitated DNA by centrifugation and resuspending it in a small volume of buffer or water (50 to 250 μL) increases the DNA concentration of the isolated DNA.

A biphasic emulsion forms when phenol and chloroform are added to the hydrophilic suspension of lysed cells (lysate). At the proper pH, centrifugation will settle the hydrophobic phase on the bottom, with the hydrophilic phase on top. Lipids and other hydrophobic components will dissolve in the lower hydrophobic phase. DNA will dissolve in the upper aqueous phase. Amphiphilic components, which have both hydrophobic and hydrophilic properties and cell debris, will collect as a white precipitate at the interface between the two layers.

The DNA in the upper phase is then precipitated by mixing with ethanol or isopropanol and salt (ammonium, potassium or sodium acetate, or lithium or sodium chloride). The ethyl or isopropyl alcohol is added to the upper-phase solution at 2:1 or 1:1 ratios, respectively, and the DNA forms a solid precipitate.

Advanced Concepts

Both ethanol and isopropanol are used for molecular applications. The ethanol is one of the general-use formulas, reagent grade. Reagent-grade alcohol (90.25% ethanol, 4.75% methanol, 5% isopropanol) is denatured; that is, the ethanol is mixed with other components because pure 100% ethanol cannot be distilled. The isopropanol used is undenatured, or pure, because it is composed of 99% isopropanol and 1% water, with no other components.

The choice of which alcohol to use depends on the starting material, the size and amount of DNA to be isolated, and the design of the method. Isopropanol is less volatile than ethanol and precipitates DNA at room temperature. Precipitation at room temperature reduces coprecipitation of salt. Also, compared with ethanol, less isopropanol is added for precipitation; therefore, isopropanol can be more practical for large-volume samples. For low concentrations of DNA, longer precipitation times at freezer temperatures may be required to maximize the amount of DNA that is recovered. An important consideration to precipitating the DNA at freezer temperatures is that the increased viscosity of the alcohol at low temperatures will require longer centrifugation times to pellet the DNA.

Recovery of minimal amounts of DNA can be optimized using carrier molecules. In early studies, yeast RNA or glycogen was used to coprecipitate low concentrations of DNA.[9] More recently, glycogen or linear polyacrylamide have been used for this purpose. Glycogen is a biological material derived from mussels (e.g., *Mytilus edulis*). As a result, some preparations of glycogen may contain DNA.[10] Commercial preparations may be treated with DNase to avoid nucleic acid contamination. Commercial glycogen may also include a color indicator to aid in detecting small pellets. Generally, 10 to 20 micrograms of carrier is added per milliliter of isopropanol mixture before incubation and centrifugation.

The DNA precipitate is collected by centrifugation. Excess salt is removed by rinsing the pelleted nucleic acid in 70% ethanol, centrifuging, and discarding the supernatant, then dissolving the DNA pellet in rehydration buffer, usually 10 mM Tris, 1 mM EDTA (TE), or water.

Inorganic Isolation Methods

Safety concerns in the clinical laboratory make the use of caustic reagents such as phenol undesirable. Methods of DNA isolation that do not require phenol extraction have therefore been developed and are used in many laboratories. Initially, these methods did not provide the efficient recovery of clean DNA achieved with phenol extraction; however, newer methods have proven to produce high-quality DNA preparations in good yields.

Inorganic DNA extraction is sometimes called **“salting out”** (Fig. 3.2). It makes use of low-pH and high-salt conditions to selectively precipitate proteins, leaving the DNA in solution. The DNA can then be precipitated as described previously using isopropanol, pelleted and resuspended in TE buffer or water.

Advanced Concepts

The presence of the chelating agent, EDTA, protects the DNA from damage by DNases present in the environment of DNA preparations intended for long-term storage. EDTA is a component of TE buffer (10 mM Tris, 1 mM EDTA) and other resuspension buffers. The EDTA will also inhibit enzyme activity when the DNA is used in various procedures, such as restriction enzyme digestion or PCR. One must be careful not to dilute the DNA too far so that large volumes (e.g., more than 10% of a reaction volume) of the DNA–EDTA solution are required for subsequent analysis methods. When DNA yield is low, as is the case with some clinical samples, it is better to dissolve it in water. More of this can be used in analysis procedures without adding excess amounts of EDTA. Because most of or the entire sample will be used for analysis, protection in storage is not a concern.

Advanced Concepts

Precipitation of the DNA excludes hydrophilic proteins, carbohydrates, and other residual contaminants still present after protein extraction. In addition, the concentration of the DNA can be controlled by adjusting the **buffer** or water volume used for resuspension of the pellet.

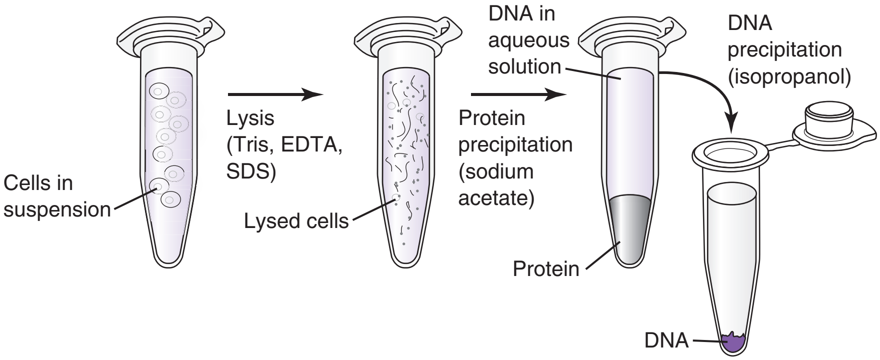


FIGURE 3.2 Inorganic DNA isolation does not include the organic extraction step. Proteins are precipitated from the cell lysate with high salt concentration and low pH. The supernatant containing the DNA is then mixed with ethanol or isopropanol to precipitate the DNA to be collected and resuspended in a smaller volume.

Solid-Phase Isolation

More rapid and comparably effective DNA extraction can be performed using solid matrices to bind and hold the DNA for washing. Silica-based products were shown to effectively bind DNA in high-salt conditions.[11] Many variations on this procedure were developed, including the use of diatomaceous earth as a source of silica particles.[12] Modern systems can be purchased with solid matrices in the form of columns or beads. Columns come in various sizes, depending on the amount of DNA to be isolated. Columns used in the clinical laboratory are most often small "spin columns" that fit inside microcentrifuge tubes. These columns are commonly used to isolate viral and bacterial DNA from serum, plasma, or cerebrospinal fluid. They are also used routinely for isolation of cellular DNA in genetics and oncology. Preparation of samples for isolation of DNA on solid-phase media starts with cell lysis and release of nucleic acids, similar to organic and inorganic procedures (Fig. 3.3). Specific buffers are used to lyse bacterial, fungal, or animal cells. Buffer systems designed for specific applications (e.g., bacterial cell lysis or human cell lysis) are commercially available.

Advanced Concepts

Alkaline lysis can be used to specifically select for plasmid DNA because chromosomal DNA will not renature properly upon neutralization and form a precipitate. The denatured chromosomal DNA and protein can be removed by centrifugation before the supernatant containing plasmid DNA is applied to the column.

Advanced Concepts

Solid matrices conjugated to specific sequences of nucleic acid can also be used to select for DNA containing complementary sequences by hybridization. After removal of noncomplementary DNA, the bound complementary DNA can be eluted by heating the matrix or by breaking the hydrogen bonds chemically.

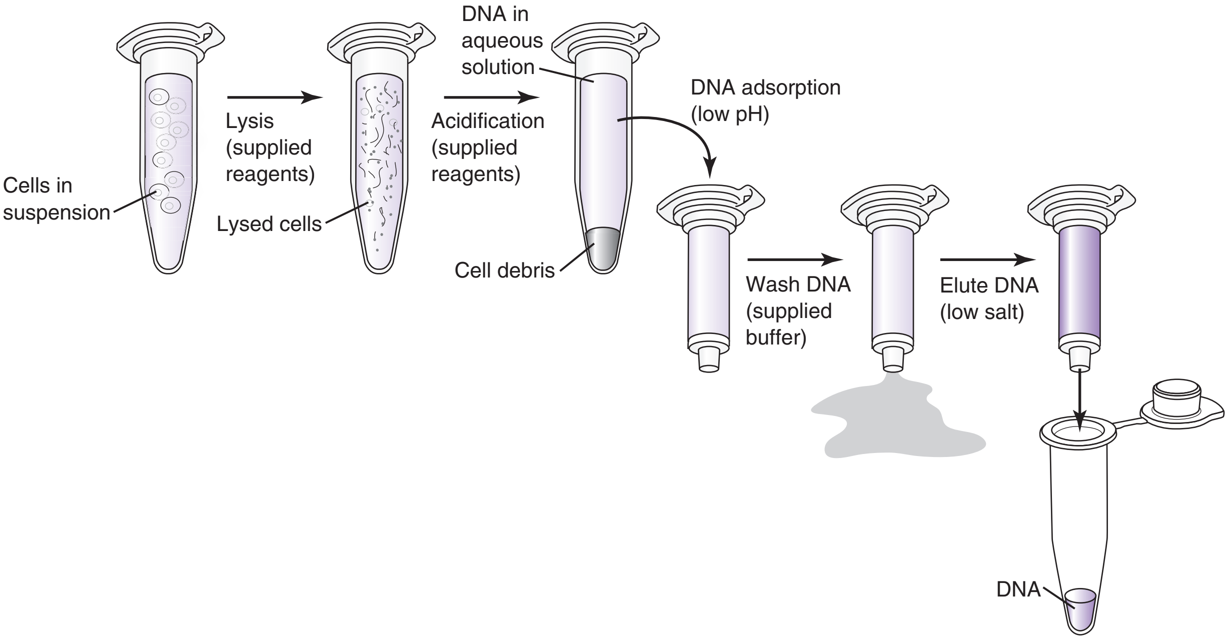


FIGURE 3.3 Isolation of DNA on solid media. Before applying to the silicate column, particulates from the cell debris may be removed by centrifugation. Solid-phase systems include the proprietary buffers and solutions for adhering the DNA on the column, washing it, and eluting it in a small volume (10 to 100 μL).

For solid-phase separation, the cell lysate is applied to a column in high-salt buffer, and the DNA in solution adsorbs to the solid matrix. Carrier RNA or DNA is applied with the DNA to enhance recovery, preventing irreversible binding of the small amount of target DNA. The carrier must be a nucleic acid of more than 200 nucleotides in length in order to bind to the silica membrane. Other carriers used to enhance precipitation, such as glycogen and linear polyacrylamide, cannot be used in this application. After the **immobilized** DNA is washed with buffer, the DNA is eluted in a specific volume of water, TE, or another low-salt buffer. The washing solutions and the eluant can be drawn through the column by gravity, vacuum, or centrifugal force. DNA absorbed to magnetic beads is washed by suspension of the beads in buffer and collection of the beads using a magnet applied to the outside of the tube while the buffer is aspirated or poured off. The DNA IQ system (Promega) uses a magnetic resin that holds a specific amount of DNA (100 ng). When the DNA is eluted in 100 µL, the DNA concentration is known (1 ng/µL) and ready for analysis.

Solid-phase isolation is the methodology employed for most robotic DNA isolation systems that use magnetized glass beads or membranes to bind DNA (Fig. 3.4). These systems are finding increased use in clinical laboratories for the automated isolation of DNA from cultures, blood, tissue, bone marrow, plasma, and other body fluids. DNA isolated from specimens, such as urine and amniotic fluid, has been reported to work better in PCR applications, possibly due to the removal of the inhibitory substances found in these types of specimens by the prefiltering step. In most systems, a measured amount of sample—for example, 200 to 1,000 µL of whole blood, 100-µL cell suspensions (10^6 to 10^7 cells), or 10 to 50 mg of tissue, in sample tubes—is placed into the instrument along with cartridges or racks of tubes containing the reagents used for isolation. Reagents are formulated in sets depending on the type and amount of starting material. The instrument is then programmed to lyse the cells and isolate and elute the DNA automatically.

Proteolytic Lysis of Fixed Material

Although high-quality DNA preparations are tantamount to successful procedures, there are circumstances that either preclude or prohibit extensive DNA purification. These include screening large numbers of samples by simple methods (e.g., electrophoresis with or without restriction enzyme digestion and some amplification procedures), isolation of DNA from limited amounts of starting material, and isolation of DNA from challenging samples such as fixed, paraffin-embedded tissues. In these cases, simple lysis of cellular material in the sample will yield sufficiently useful DNA for amplification procedures.

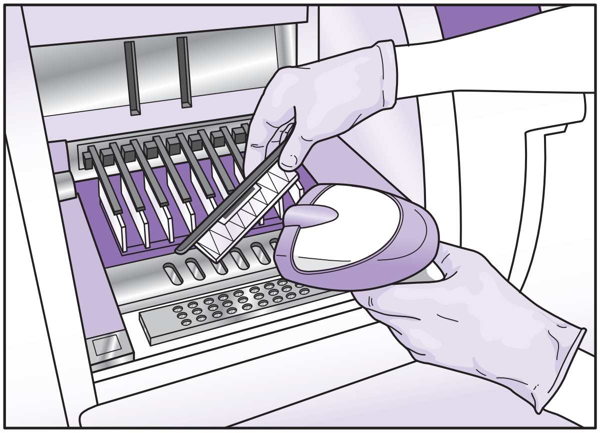


FIGURE 3.4 Automated DNA isolation systems use cartridges containing lysis, adsorption, and elution agents and magnetic beads. Bar coding of reagents and sample tubes assures accurate sample and reagent lot tracking.

Simple screening methods are mostly used for research purposes in which large numbers of samples must be processed. In contrast, analysis of paraffin samples is frequently performed in the clinical laboratory. Fixed tissue is more conveniently accessed in the laboratory and may sometimes be the only source of patient material. Thin sections are usually used for analysis, although sectioning is not necessary with very small samples such as needle biopsies. Paraffin-embedded specimens can be dewaxed with xylene or other agents and then rehydrated before nucleic acid isolation, the fixed tissue can be used without dewaxing. For some tests, such as somatic mutation analyses, an adjacent stained serial section can be examined microscopically to identify areas of the tissue comprising the appropriate percentage of tumor cells. The identifiable areas of the tumor can then be isolated directly

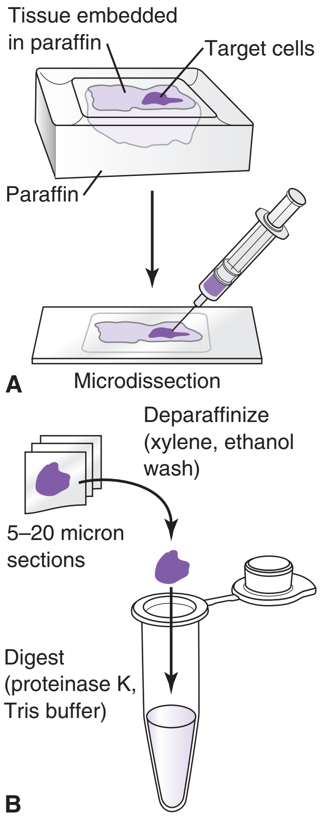

Tissue embedded in paffin
Target cells
Paraffin
A
Microdissection
Deparaffinize (xylene, ethanol wash)
5–20 micron sections
Digest (proteinase K, Tris buffer)
B

FIGURE 3.5 Crude extraction of DNA from fixed, paraffin-embedded tissue. Selected regions of tissue are scraped from slides (macrodissection; A) and extracted in proteinase K (B).

from the slide by lifting or scraping in buffer with or without the use of a microscope (microdissection or macrodissection, respectively; Fig. 3.5) or laser capture microscopy.

Before lysis, cells may be washed by suspension and centrifugation in saline or other isotonic buffers. Reagents used for cell lysis depend on the subsequent use of the DNA. For simple screens, cells can be lysed in detergents, such as SDS or Triton. For use in PCR amplification, cells may be lysed in a mixture of Tris buffer and proteinase K. The proteinase K will digest proteins in the sample, lysing the cells and inactivating other enzymes. The released DNA can be used directly in the amplification reaction.

Rapid Extraction Methods

With the advent of PCR, new and faster methods of DNA isolation have been developed. The minimal sample requirements of amplification procedures allow for the use of material previously not utilizable for analysis. Rapid lysis methods and DNA extraction/storage cards provide sufficiently clean DNA that can be used for amplification.

Chelex is a cation-chelating resin that can be used for simple extraction of DNA from minimal samples.[13] A suspension of 10% Chelex resin beads is mixed with the specimen, and the cells are lysed by boiling. After centrifugation of the suspension, DNA in the supernatant is cooled and may be further extracted with chloroform before use in amplification procedures. This method is most commonly used in forensic applications[14,15] but may also be useful for purification of DNA from clinical samples and fixed, paraffin-embedded specimens[16] and samples spotted on storage cards.[17,18]

Isolation of Mitochondrial DNA

There are two approaches to the isolation of mitochondrial DNA from eukaryotic cells. One method is to first isolate the mitochondria by centrifugation. After cell preparations are homogenized by grinding on ice, the homogenate is centrifuged at low speed (700 to 2,600 × *g*) to pellet intact cells, nuclei, and cell debris. The mitochondria are pelleted from the supernatant in a second high-speed centrifugation (10,000 to 16,000 × *g*) and lysed with detergent. The lysate is treated with proteinase to remove protein contaminants. Mitochondrial DNA can then be precipitated with cold ethanol and resuspended in water or appropriate buffers for analysis. The second approach to mitochondrial DNA preparation is to isolate total DNA as described previously. The preparation will contain mitochondrial DNA that can be analyzed within the total DNA background by hybridization or PCR.

Advanced Concepts

When homogenizing cells for isolation of mitochondria, care must be taken not to overgrind the tissue and dissociate the mitochondrial membranes.

Grinding in alkaline buffers with reducing agents such as β-mercaptoethanol will protect the mitochondria during the isolation process. A high-ionic-strength buffer can also be used to selectively lyse the nuclear membranes.

ISOLATION OF RNA

Working with RNA in the laboratory requires strict precautions to avoid sample RNA degradation. RNA is especially labile due to the ubiquitous presence of RNases. These enzymes are small proteins that can renature, even after autoclaving, and regain activity. They remain active at a wide range of temperatures (down to –20°C and below). Unlike DNases, RNases must be eliminated or inactivated before isolation of RNA.

It is necessary to allocate a separate **RNase-free (RNF)** area of the laboratory for storage of materials and specimen handling intended for RNA analysis. Gloves must always be worn in the RNF area. Disposables (tubes, tips, etc.) that come in contact with the RNA should be kept at this location and never be touched by ungloved hands. Articles designated DNAse-free/RNF by suppliers may be used directly from the package. Although reusable glassware is seldom used for RNA work, it can be rendered RNF. After cleaning, glassware must be baked for four to six hours at 400°C to inactivate the RNases.

Total RNA

There are several types of RNA in prokaryotes and eukaryotes. The most abundant (80% to 90%) RNA in all cells is ribosomal RNA (rRNA). This RNA consists of two components, large and small, which are visualized by agarose gel electrophoresis (see Fig. 3.10). Depending on the cell type and conditions, the next most abundant RNA fraction (2.5% to 5%) is messenger RNA (mRNA). This mRNA may be detected as a faint background underlying the rRNA detected by agarose gel electrophoresis. Transfer RNA and small nuclear RNAs are also present in the total RNA sample.

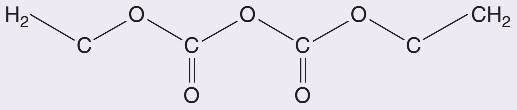
Specimen Collection

RNA tests, especially those that involve gene expression, such as array analysis or quantitative reverse transcriptase PCR, are most accurate when the RNA is stabilized from further metabolism or degradation after collection. Specialized collection tubes have been devised to stabilize RNA immediately upon blood draw. In the laboratory, tissue or cell pellets can be stored and/or transported in preservative reagents.[13,19,20]

Advanced Concepts

Several chemical methods have been developed to inactivate or eliminate RNases. **Diethyl pyrocarbonate (DEPC)** can be added to water and buffers (except for Tris buffer) to inactivate RNases permanently. DEPC converts primary and secondary amines to carbamic acid esters. It cross-links RNase proteins through intermolecular covalent bonds, rendering them insoluble. Because of its effect on amine groups, DEPC will adversely affect Tris buffers. DEPC will also interact with polystyrene and polycarbonate and is not recommended for use on these types of materials.

Other RNase inhibitors include vanadyl-ribonucleoside complexes, which bind the active sites of the RNase enzymes, and macaloid clays, which absorb the RNase proteins. Ribonuclease inhibitor proteins can be added directly to RNA preparations. These proteins form a stable noncovalent complex with the RNases in solution. Some of these interactions require reducing conditions; therefore, dithiothreitol might be added in addition to the inhibitor.



Diethyl pyrocarbonate.

RNA Isolation Chemistries

Preparation of specimen material for RNA extraction differs in some respects from DNA extraction. Reticulocytes in blood and bone marrow samples are lysed by osmosis or separated from WBCs by centrifugation. When dissociating tissue, the sample should be kept frozen in liquid nitrogen or immersed in buffer that will inactivate **intracellular** RNases. This is especially true for tissues such as pancreas that contain large amounts of innate RNases. Bacterial and fungal RNA are also isolated by chemical lysis or by grinding in liquid nitrogen. Viral RNA can be isolated directly from serum, plasma, culture medium, or other cell-free fluids by means of specially formulated spin columns or beads. Cell-free material is used for these isolations because most total RNA isolation methods cannot distinguish between RNA from microorganisms and those from host cells.

Organic Isolation

The cell lysis step for RNA isolation is performed in detergent or phenol in the presence of high salt (0.2 to 0.5 M NaCl) or RNase inhibitors. Guanidine isothiocyanate (GITC) is a strong denaturant of RNases and can be used instead of high-salt buffers. Strong reducing agents such as 2-mercaptoethanol may also be added during this step.

Once the cells are lysed, proteins can be extracted by organic means (Fig. 3.6). Acid phenol:chloroform:isoamyl alcohol (25:24:1) solution efficiently extracts RNA. Chloroform enhances the extraction of the nucleic acid by denaturing proteins and promoting phase separation. Isoamyl alcohol prevents foaming. For RNA, the organic phase must be acidic (pH 4 to 6). The acidity of the organic phase is adjusted by overlaying it with a buffer of the appropriate pH. In some isolation procedures, DNase is added at the lysis step to eliminate contamination of DNA. Alternatively, RNF DNase also may be added directly to the isolated RNA at the end of the procedure. After phase separation, the upper aqueous phase containing the RNA is removed to a clean tube, and the RNA is precipitated by addition of two volumes of ethanol or one volume of isopropanol. Glycogen or yeast-transfer RNA may be added at this step as a carrier to aid RNA pellet formation. The RNA precipitate is then washed in 70% ethanol and resuspended in RNF buffer or water.

Solid-Phase Isolation

Solid-phase separation of RNA begins with similar steps as described previously for organic extraction. The strong denaturing buffer conditions must be adjusted before application of the lysate to the column (Fig. 3.7). In some procedures, ethanol is added at this point. Some systems provide a filter column to remove particulate material before application to the adsorption column. As with DNA columns, commercial reagents are supplied with the columns to optimize RNA adsorption and washing on the silica-based matrix.

After lysate is applied to the column in the high-salt chaotropic buffer, the adsorbed RNA is washed with the supplied buffers. DNase may be added directly to the adsorbed RNA on the column to remove contaminating DNA. Washing solutions and the eluant are drawn through the column by gravity, vacuum, or centrifugal force. Small columns of silica-based material that fit inside **microfuge** tubes (spin columns) are widely

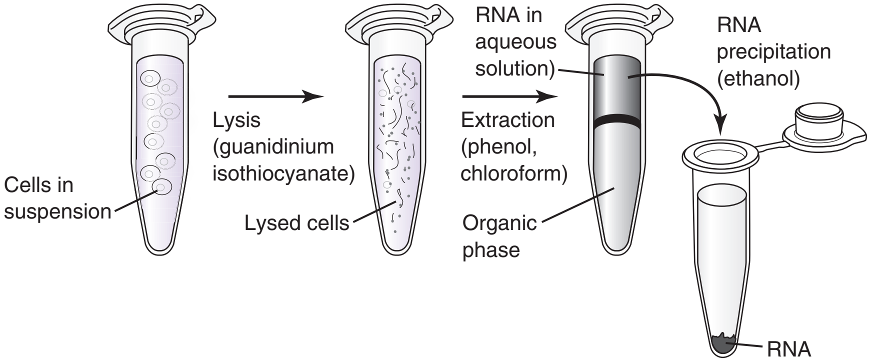

Cells in suspension
Lysis (guanidinium isothiocyanate)
Lysed cells
RNA in aqueous solution)
Extraction (phenol, chloroform)
Organic phase
RNA precipitation (ethanol)
RNA

FIGURE 3.6 Organic extraction of total RNA is similar to that of DNA; however, the RNA must be protected from intracellular and extracellular RNases. The RNA in solution is precipitated and resuspended, similar to DNA. Some procedures include treatment of the RNA solution with DNase to remove any contaminating DNA.

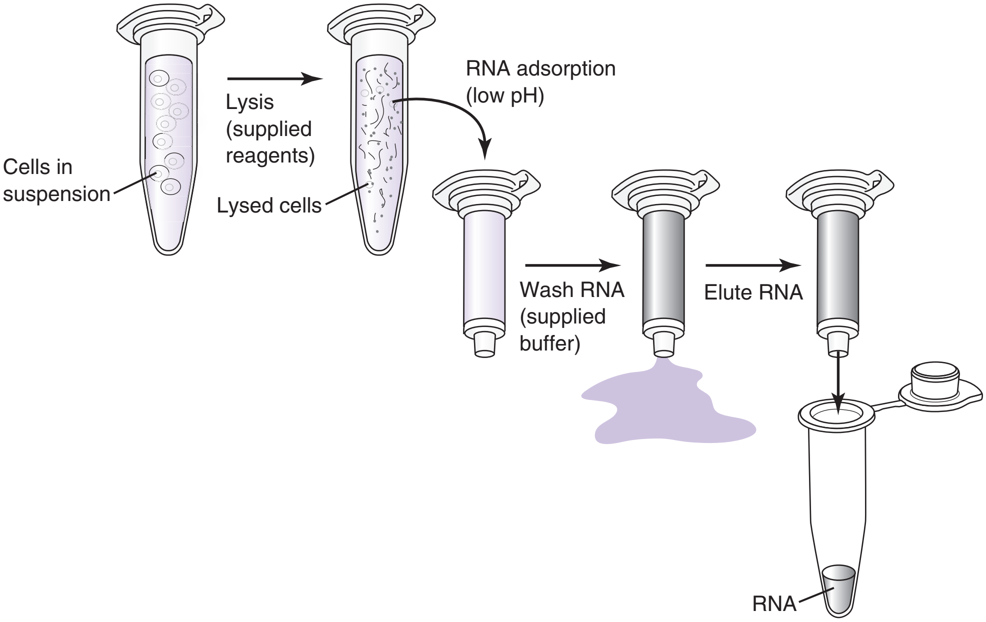


FIGURE 3.7 Isolation of RNA on a solid matrix is also similar to DNA isolation. Buffers and wash solutions are designed for the RNA target.

used for routine nucleic acid isolation from all types of specimens. The eluted RNA is usually of sufficient concentration and purity for direct use in most applications. Automated systems like those shown in Figure 3.4 use magnetic beads or particles for RNA isolation as well. Special reagent sets are available for RNA or total nucleic acid (RNA and DNA) on these instruments.

Generally, 1 million eukaryotic cells or 10 to 50 mg of tissue will yield about 10 μg of RNA. The yield of RNA from biological fluids will depend on the concentration of microorganisms or other target molecules present in the specimen (Table 3.3).

Proteolytic Lysis of Fixed Material

The quality of RNA from fixed, paraffin-embedded tissue will depend on the fixation process and handling of the specimen before fixation (Table 3.2).[21,22] Commercial reagent sets are available for extraction of RNA from fixed-tissue specimens. These reagents are designed to provide lysis and incubation conditions that reverse formaldehyde modification of RNA and efficiently release RNA from tissue sections while avoiding further RNA degradation. Specialized reagents or spin columns for removal of genomic DNA contamination are included in some systems. Automated isolation systems also have reagent kits and cartridges designed to isolate RNA from fixed tissue.

Isolation of polyA (Messenger) RNA

As previously stated, approximately 80% to 90% of total RNA is rRNA. Often the RNA required for analysis is mRNA, accounting for only about 2.5% to 5% of the total RNA yield. The majority of mRNA consists of mRNA from highly expressed genes. Rare or single-copy mRNA is, therefore, a very minor part of the total RNA isolation. To enrich the yield of mRNA, especially rare transcripts, protocols employing single-stranded oligomers of thymine or uracil immobilized on a matrix resin column or beads are used (Fig. 3.8). The polyT or polyU oligomers will bind the polyA tail found exclusively on mRNA. After washing away residual RNA, polyA RNA is eluted by washing the column with warmed, low-salt buffer containing detergent to break the hydrogen bonds between the mRNA and the column. With this approach, approximately 30 to 40 ng of mRNA can be obtained from 1 μg of total RNA.

TABLE 3.3 Yield of RNA From Various Specimen Sources[36-38]

Specimen	Expected Yield*
Blood† (1 mL, 3.5–10 × 10^6 WBCs/mL)	1–10 µg
Buffy coat† (1 mL whole blood)	5–10 µg
Bone marrow† (1 mL)	50–200 µg
Cultured cells‡ (10^7 cells)	50–150 µg
Buccal cells (1 mg)	1–10 µg
Solid tissue§ (1 mg)	0.5–4 µg
Fixed tissue‖ (1 mm^3)	0.2–3 µg
Bacterial culture¶ (0.5 mL, 0.7 absorbance units)	10–100 µg

*Specimen handling especially affects RNA yield. Isolation of polyA RNA will result in much lower yields. See text.
†RNA yield will depend on WBC count.
‡RNA yield will depend on type of cells and the conditions of cell culture.
§Liver, spleen, and heart tissues yield more RNA than brain, lung, ovary, kidney, or thymus tissues.
‖Isolation of RNA from fixed tissue is especially affected by the type of fixative used and the age and the preliminary handling of the original specimen.
¶Different bacterial types and fungi will yield more or less RNA.

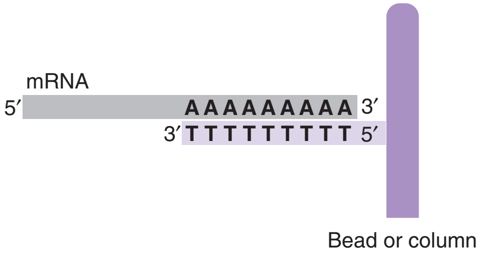


FIGURE 3.8 Oligo polythymine (or polyuracil) columns or beads bind the polyA tail of mRNA. Peptide nucleic acid dT or dU can also be used. These polymers are less subject to degradation by contaminating nucleases.

There are limitations to the isolation of polyA RNA using oligo dT or dU. The efficiency of polyA and polyU binding is variable. Secondary structure (intrastrand or interstrand hydrogen bonds) in the target sample may compete with binding to the capture oligomer. Also, mRNAs with short polyA tails may not bind efficiently or at all. AT-rich DNA fragments might bind to the column and not only compete with the desired mRNA target but also contaminate the final eluant. Potential digestion of the oligo-conjugated matrices precludes the use of DNase on the RNA before it is bound to the column. Treatment of the eluant with RNF DNase is possible, but the DNase should be inactivated if the mRNA is to be used in procedures involving DNA components. Furthermore, rRNA may co-purify with the polyA RNA. The final product, then, is enriched in polyA RNA but is not a pure polyA preparation.

MEASUREMENT OF NUCLEIC ACID QUALITY AND QUANTITY

Laboratory analysis of nucleic acids produces variable results, depending on the quality and quantity of input material. This is an important consideration in the medical laboratory because test results must be accurately interpreted with respect to disease pathology. Consistent results require that run-to-run variation be minimized. Fortunately, measurement of the quality and quantity of DNA and RNA is straightforward.

Electrophoresis

DNA and RNA can be analyzed for quality by resolving an aliquot of the isolated sample on an agarose gel (Fig. 3.9). Fluorescent dyes such as ethidium bromide or SybrGreen I bind specifically to DNA and are used to visualize the sample preparation. Ethidium bromide or SybrGreen II are used to detect RNA. Less frequently, **silver stain** has been used to detect small amounts of DNA by visual inspection.

The appearance of DNA on agarose gels depends on the type of DNA isolated. A good preparation of plasmid DNA will yield a bright signal from supercoiled plasmid DNA with minor or no other bands that represent nicked or broken plasmid (see Fig. 3.10). High-molecular-weight chromosomal DNA should collect as a bright band near the top of the gel. A high-quality preparation of RNA will yield two distinct bands of ribosomal RNA. The integrity of these bands is an indication of the integrity of the other RNA species present in the same sample. If these bands are degraded (smeared) or absent,

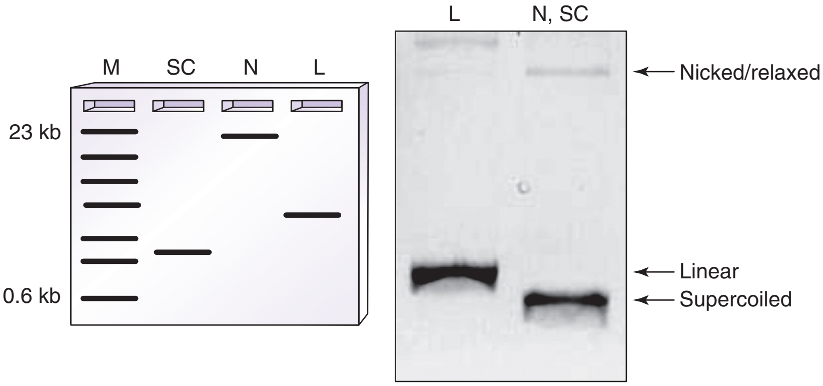


FIGURE 3.9 After agarose gel electrophoresis, compact supercoiled plasmid DNA (SC) will travel farther through the gel than nicked plasmid (N), which has single-strand breaks. Relaxed plasmid DNA (R) has double-strand breaks and will migrate according to its size, 23 kb in the drawing on the left. Linear (L) plasmids migrate according to the size of the plasmid. A gel photo shows a plasmid preparation. *N,* nicked; *SC,* supercoiled; *L,* linear; *R,* relaxed; *M,* molecular-weight markers.

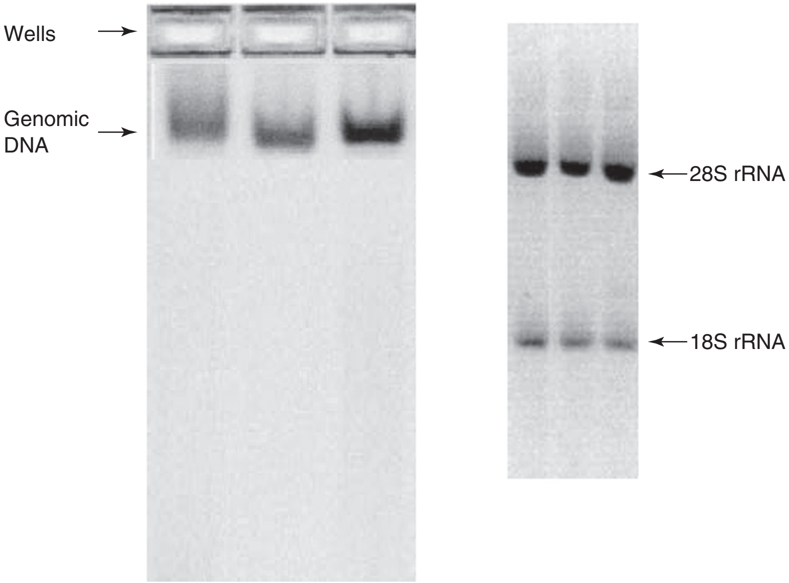


FIGURE 3.10 Intact ethidium bromide–stained human chromosomal DNA *(left)* and total RNA *(right)* after agarose gel electrophoresis. High-quality genomic DNA runs as a tight smear close to the loading wells. High-quality total RNA appears as two rRNA bands representing large and small ribosomal RNA species (shown with molecular weight markers, *M*).

the quality of the RNA in the sample is deemed unacceptable for use in molecular assays.

When fluorescent dyes are used, DNA and, less accurately, RNA can be quantified by comparison of the fluorescence intensity of the sample aliquot run on the gel with that of a known amount of control DNA or RNA loaded on the same gel. Densitometry of the band intensities gives the most accurate measurement of quantity. For some procedures, estimation of DNA or RNA quantity can be made by visual inspection.

Spectrophotometry

Nucleic acids absorb light at 260 nm through the adenine residues. Using the Beer–Lambert law, concentration can be determined from the **absorptivity constants** (50 for double-stranded DNA, 40 for RNA). The relationship of concentration to absorbance is expressed as

$$A = \in bc$$

where A = absorbance, ∈ = molar absorptivity (L/mol-cm), b = path length (cm), and c = concentration (mg/L). The absorbance at this wavelength is thus directly proportional to the concentration of the nucleic acid in the sample. Using the absorptivity as a conversion factor from optical density to concentration, one optical density unit (or absorbance unit) at 260 nm is equivalent to 50 mg/L (or 50 µg/mL) of double-stranded DNA and 40 µg/mL of RNA. To determine concentration, multiply the spectrophotometer reading in absorbance units by the appropriate conversion factor. Phenol absorbs ultraviolet light at 270 to 275 nm, close to the wavelength of maximum absorption by nucleic acids. This means that residual phenol from organic isolation procedures can increase 260 readings, so phenol contamination must be avoided when measuring concentration at 260 nm.

If the DNA or RNA preparation is of sufficient concentration to require dilution before spectrophotometry in order for the reading to fall within the linear reading range, the dilution factor must be included in the calculation of quantity. Multiply the absorbance reading by the conversion factor and the dilution factor to find the concentration of nucleic acid.

EXAMPLE 1: A DNA preparation diluted $^1/_{100}$ yields an absorbance reading of 0.200 at 260 nm. To obtain the concentration in µg/mL, multiply:

$$0.200 \text{ absorbance units} \times 50 \ \mu\text{g/mL} \text{ per absorbance unit} \times 100 = 1{,}000 \ \mu\text{g/mL}$$

The yield of the sample is calculated using the volume of the preparation. If the DNA was eluted or resuspended in a volume of 0.5 mL in the case illustrated previously, the yield would be

$$1{,}000 \ \mu\text{g/mL} \times 0.5 \text{ mL} = 500 \ \mu\text{g}$$

EXAMPLE 2: An RNA preparation diluted $^1/_{10}$ yields an absorbance reading of 0.500 at 260 nm. The concentration is

$$0.500 \text{ absorbance units} \times 40 \ \mu\text{g/mL} \text{ per absorbance unit} \times 10 = 200 \ \mu\text{g/mL}$$

The yield of the sample is calculated using the volume of the preparation. If the DNA was eluted or resuspended in 0.2 mL in the case illustrated previously, the yield would be

$$200 \ \mu\text{g/mL} \times 0.2 \text{ mL} = 40 \ \mu\text{g}$$

Historical Highlights

The maximum absorption of light at 260 nm was one of the clues suggesting that DNA was the molecular matter of genetic material. In 1942, Lewis Stadler reported results on studies of the effect of wavelength of ultraviolet (UV) light on mutagenesis in corn plants.[23] He observed that the most mutagenic monochromatic light had a wavelength of 260 nm. Together with the data of Avery[24] and observations of Hershey and Chase,[25] Stadler's observations further supported the idea that DNA is the carrier of genetic information.

Spectrophotometric measurements may also be used to estimate the purity of nucleic acid. Ideally, contamination can be detected by reading the concentration over a range of wavelengths. Graphing the absorbance reading as a function of wavelength produces a curve or spectrum that should peak at A_{260} nm for nucleic acid. Some spectrophotometers produce this graph automatically. Even with such automation, however, spectral analysis for each specimen sample is not practical for most routine medical laboratory work. In practice, nucleic acid solutions are read at two or three distinct wavelengths (Table 3.4). Absorbance over background at any wavelength other than the A_{260} maxima of the nucleic acid indicates contamination.

TABLE 3.4 Common Contaminants and Their Wavelengths of Peak Absorbance

Wavelength (nm)	Contaminant
230	Organic compounds
270	Phenol
280	Protein
>330	Particulate matter

Routine nucleic acid isolation procedures produce reasonably pure DNA solutions free from particles, organic compounds, and phenol. Due to its abundance and close association with nucleic acid in the cell, the most likely contaminant in a nucleic acid preparation will be protein. Protein absorbs light at 280 nm through the aromatic tryptophan and tyrosine residues. Although this measurement is not highly accurate, general ranges indicate at least the presence of protein contaminants. The absorbance of the nucleic acid at 260 nm should be 1.6 to 2.00 times more than the absorbance at 280 nm. If the 260-nm/280-nm ratio is less than 1.6, the nucleic acid preparation may be contaminated with unacceptable amounts of protein and not of sufficient purity for use. Such a sample can be improved by column purification, reprecipitating the nucleic acid, or repeating the protein-removal step of the isolation procedure. It should be noted that low pH affects the 260-nm/280-nm ratio. Somewhat alkaline buffers (pH 7.5) are recommended for accurate determination of purity. RNA affords a somewhat higher 260-nm/280-nm ratio, 2.0 to 2.3. A DNA preparation with a ratio higher than 2.0 may be contaminated with RNA. Some procedures for DNA analysis are not affected by contaminating RNA, in which case the DNA is still suitable for use. If, however, RNA may interfere or react with DNA-detection components, RNase should be used to remove the contaminating RNA. Because it is difficult to detect contaminating DNA in RNA preparations, RNA should be treated with RNF DNase where DNA contamination may interfere with subsequent applications.

Absorbance ratios do not necessarily predict successful amplification or hybridization reactions. Because the light is absorbed by single-nitrogen-base moieties, fragmented or degraded DNA will produce an A_{260} reading comparable to intact DNA. In this regard, fluorometry is considered by some to be a better method of quality assessment.

Historical Highlights

Estimation of protein contamination in nucleic acid is based on a method by Warburg and Christian designed to measure nucleic acid contamination in protein.[26] This method utilizes the natural light absorption at 280 nm of the aromatic amino acid side chains, tryptophan, and tyrosine. Nucleic acids are common contaminants of protein preparations and naturally absorb light at 260 nm, where there is no interference by protein. In the Warburg–Christian method, the test material was read at 260 and 280 nm, and then a nomograph was used to determine the relative amounts of protein and nucleic acid.

Advanced Concepts

Ultraviolet spectrophotometers dedicated to nucleic acid analysis can be programmed to calculate absorbance ratios and concentrations. The operator must enter the type of nucleic acid, the dilution factor, and the desired conversion factor. The instrument will automatically read the sample at the appropriate wavelengths and do the required calculations, giving a reading of concentration in µg/mL and a 260-nm/280-nm ratio.

Fluorometry

Fluorometry, or fluorescent spectroscopy, measures fluorescence related to DNA concentration in association with DNA-specific fluorescent dyes. Early methods used 3,5-diaminobenzoic acid 2HCl (DABA).[27] This dye combines with alpha-methylene aldehydes (deoxyribose) to yield a fluorescent product. It is still used for the detection of nuclei and as a control for hybridization and spot integrity in microarray analyses. Although less convenient to use, fluorometry is recommended for procedures where accurate measurement of intact DNA is critical, such as next-generation sequencing.

More modern procedures use the DNA-specific dye **Hoechst 33258.** This dye combines with adenine-thymine base pairs in the minor groove of the DNA double helix and is thus specific for intact double-stranded DNA. The binding specificity for A-T residues, however, complicates measurements of DNA that have unusually high or low GC content. A standard measurement is required to determine concentration by fluorometry, and this standard must have GC content similar to that of the samples being measured. Calf thymus DNA (GC content = 50%) is often used as a standard for specimens with unknown DNA GC content. Fluorometric determination of DNA concentration using Hoechst dye is very sensitive (down to 200 ng DNA/mL).

PicoGreen and OliGreen are other DNA-specific dyes that can be used for fluorometric quantification. Due to brighter fluorescence upon binding to double-stranded DNA, PicoGreen is more sensitive than Hoechst dye (detection down to 25 pg/mL concentrations). Like Hoechst dye, single-stranded DNA and RNA do not bind to PicoGreen. OliGreen is designed to bind to short pieces of single-stranded DNA (oligonucleotides). This dye will detect down to 100 pg/mL of single-stranded DNA. OliGreen will not fluoresce when bound to double-stranded DNA or RNA.

RNA may be measured in solution using SybrGreen II RNA gel stain.[28] The intensity of SyBrGreen II fluorescence is 20% to 26% lower with polyadenylated RNA than with total RNA. The sensitivity of this dye is 2 ng/mL. SybrGreen II, however, is not specific to RNA and will bind and fluoresce with double-stranded DNA as well. Contaminating DNA must therefore be removed in order to get an accurate determination of RNA concentration.

Fluorometry measurements require calibration of the instrument with a known amount of standard before measurement of the sample. For methods using Hoechst dye, the dye, diluted to a working concentration of 1 µg/mL in water, is mixed with the sample (usually a dilution of the sample). Once the dye and sample solution are mixed, fluorescence must be read within two hours because the dye/sample complex is stable only for approximately this amount of time. The fluorescence is read in a quartz cuvette. A programmed fluorometer will read out a concentration based on the known standard calibration.

Absorption and fluorometry readings may not always agree. First, the two detection methods recognize different targets. Single nucleotides do not bind to fluorescent dyes, but they can absorb ultraviolet light and affect spectrometric readings. Furthermore, absorption measurements do not distinguish between DNA and RNA, so contaminating RNA may be factored into the DNA measurement. RNA does not enhance the fluorescence of the fluorescent dyes and is thus invisible to fluorometric detection. In fact, specific detection of RNA in the presence of DNA in solution is not yet possible. Deciding which instrument to use is at the discretion of the laboratory. Most laboratories use spectrophotometry because the samples can be read directly without staining or mixing with dye. For methods that require accurate measurements of low amounts of DNA or RNA (in the range of 10 to 100 ng/mL), fluorometry may be preferred.

Microfluidics

Lab-on-a-chip technology has been applied to nucleic acid quantification and analysis using microfluidics-based instruments. To make a measurement, the sample is applied to a multiwell chip. The sample then moves through microchannels across a detector. The instrument software generates images in electropherogram (peak) or gel (band) configurations. For RNA, a quantification estimate called the RNA integrity number is determined as a standard measure of RNA integrity. The presence of ribosomal impurities in mRNA preparations may also be calculated. This system is more automated than standard spectrophotometry or fluorometry, uses a minimal volume of sample (as low as 1 µL), and can test multiple samples simultaneously. Microfluidics is useful for the analysis of studies on small RNAs, such as microRNAs in eukaryotes and gene expression in bacteria.

STUDY QUESTIONS

DNA Quantity/Quality

1. Calculate the DNA concentration in µg/mL from the following information:

a. Absorbance reading at 260 nm from a 1:100 dilution = 0.307
b. Absorbance reading at 260 nm from a 1:50 dilution = 0.307
c. Absorbance reading at 260 nm from a 1:100 dilution = 0.172
d. Absorbance reading at 260 nm from a 1:100 dilution = 0.088

2. If the volume of the DNA solutions in question 1 was 0.5 mL, calculate the yield for (a)–(d).

3. Three DNA preparations have the following A_{260} and A_{280} readings:

Sample	OD_{260}	OD_{280}
(i) 1	0.419	0.230
(ii) 2	0.258	0.225
(iii) 3	0.398	0.174

For each sample, based on the A_{260}/A_{280} ratio, is each preparation suitable for further use? If not, what is contaminating the DNA?

4. After agarose gel electrophoresis, a 0.5-microgram aliquot of DNA isolated from a bacterial culture produced only a faint smear at the bottom of the gel lane. Is this an acceptable DNA sample?

5. Compare and contrast the measurement of DNA concentration by spectrophotometry with analysis by fluorometry with regard to staining requirements and accuracy.

RNA Quantity/Quality

1. Calculate the RNA concentration in μg/mL from the following information:

   a. Absorbance reading at 260 nm from a 1:100 dilution = 0.307
   b. Absorbance reading at 260 nm from a 1:50 dilution = 0.307
   c. Absorbance reading at 260 nm from a 1:100 dilution = 0.172
   d. Absorbance reading at 260 nm from a 1:100 dilution = 0.088

2. If the volume of the RNA solutions in question 1 was 0.5 mL, calculate the yield for (a)–(d).

3. An RNA preparation has the following absorbance readings:

   $A_{260} = 0.208$
   $A_{280} = 0.096$

   Is this RNA preparation satisfactory for use?

4. A blood sample was held at room temperature for 5 days before being processed for RNA isolation. Will this sample likely yield optimal RNA?

5. Name three factors that will affect the yield of RNA from a paraffin-embedded tissue sample.

References

1. Mirsky A. The discovery of DNA. *Scientific American* 1968;218: 78–88.
2. Meselson M, Stahl FW. The replication of DNA in *Escherichia coli*. *Proceedings of the National Academy of Sciences* 1958;44: 671–682.
3. Siravegna G, Marsoni S, Siena S, Bardelli A. Integrating liquid biopsies into the management of cancer. *Nature Reviews: Clinical Oncology* 2017;14(9):531–548.
4. Grskovic M, Hiller DJ, Eubank LA, Sninsky JJ, Christopherson C, Collins JP, Thompson K, Song M, Wang YS, Ross D, Nelles MJ, Yee JP, Wilber JC, Crespo-Leiro MG, Scott SL, Woodward RN. Validation of a clinical-grade assay to measure donor-derived cell-free DNA in solid organ transplant recipients. *Journal of Molecular Diagnostics* 2016;18:890–902.
5. Thakur BK, Zhang H, Becker A, Matei I, Huang Y, Costa-Silva B, Zheng Y, Hoshino A, Brazier H, Xiang J, Williams C, Rodriguez-Barrueco R, Silva JM, Zhang W, Hearn S, Elemento O, Paknejad N, Manova-Todorova K, Welte K, Bromberg J, Peinado H, Lyden D. Double-stranded DNA in exosomes: a novel biomarker in cancer detection. *Cell Research* 2014;24:766–769.
6. Lucena-Aguilar G, Sánchez-López AM, Barberán-Aceituno C, Carrillo-Ávila JA, López-Guerrero JA, Aguilar-Quesada R. DNA source selection for downstream applications based on DNA quality indicators analysis. *Biopreservation and Biobanking* 2016;14: 264–270.
7. Greer C, Peterson SL, Kivat NB, Manos MM. PCR amplification from paraffin-embedded tissues. *American Journal of Clinical Pathology* 1991;95:117–124.
8. Perry C, Chung JY, Ylaya K, Choi CH, Simpson A, Matsumoto KT, Smith WA, Hewitt SM. A buffered alcohol-based fixative for histomorphologic and molecular applications. *Journal of Histochemistry and Cytochemistry* 2016;64:425–440.

9. Gallagher M, Burke WF Jr, Orzech K. Carrier RNA enhancement of recovery of DNA from dilute solutions. *Biochemical Biophysical Research Communications* 1987;144:271–276.
10. Bartram A, Poon C, Neufeld JD. Nucleic acid contamination of glycogen used in nucleic acid precipitation and assessment of linear polyacrylamide as an alternative co-precipitant. *BioTechniques* 2009;47:1019–1022.
11. Vogelstein B, Gillespie D. Preparative and analytical purification of DNA from agarose. *Proceedings of the National Academy of Sciences* 1979;76:615–619.
12. Carter MJ, Milton ID. An inexpensive and simple method for DNA purifications on silica particles. *Nucleic Acids Research* 1993;21:1044.
13. Altayari W. DNA extraction: organic and solid-phase. *Methods in Molecular Biology* 2016;1420:55–68.
14. Eychner A, Schott KM, Elkins KM. Assessing DNA recovery from chewing gum. *Medicine, Science and the Law* 2017;57:7–11.
15. Forsberg C, Jansson L, Ansell R, Hedman J. High-throughput DNA extraction of forensic adhesive tapes. *Forensic Science International, Genetics* 2016;24:158–163.
16. Wang J, Gouda-Vossos A, Dzamko N, Halliday G, Huang Y. DNA extraction from fresh-frozen and formalin-fixed, paraffin-embedded human brain tissue. *Neuroscience Bulletin* 2013;29:649–654.
17. Hailemariam Z, Ahmed JS, Clausen PH, Nijhof AM. A comparison of DNA extraction protocols from blood spotted on FTA cards for the detection of tick-borne pathogens by reverse line blot hybridization. *Ticks and Tick Borne Diseases* 2017;8:185–189.
18. Rahikainen A, Palo JU, de Leeuw W, Budowle B, Sajantila A. DNA quality and quantity from up to 16 years old post-mortem blood stored on FTA cards. *Forensic Science International* 2016;261:148–153.
19. Franken C, Remy S, Lambrechts N, Hollanders K, Den Hond E, Schoeters G. Peripheral blood collection: the first step towards gene expression profiling. *Biomarkers* 2016;21:458–465.
20. Duale N, Brunborg G, Rønningen KS, Briese T, Aarem J, Aas KK, Magnus P, Stoltenberg C, Susser E, Lipkin WI. Human blood RNA stabilization in samples collected and transported for a large biobank. *BMC Res Notes* 2012;5:51–59.
21. Mizumo T, Nagamura H, Iwamoto KS, et al. RNA from decades-old archival tissue blocks for retrospective studies. *Diagnostic Molecular Pathology* 1998;7:202–208.
22. Howe K. Extraction of miRNAs from formalin-fixed paraffin-embedded (FFPE) tissues. *Methods in Molecular Biology* 2017;1509:17–24.
23. Stadler L, Uber FM. Genetic effects of ultraviolet radiation in maize. IV. Comparison of monochromatic radiations. *Genetics* 1942;27:84–118.
24. Avery O, MacLeod CM, McCarty M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types. I. Induction of transformation by a DNA fraction isolated from pneumococcal type III. *Journal of Experimental Medicine* 1944;79:137–158.
25. Hershey AD, Chase M. Independent function of viral protein and nucleic acid in growth of bacteriophage. *Journal of General Physiology* 1952;26:36–56.
26. Glasel J. Validity of nucleic acid purities monitored by A260/A280 absorbance ratios. *Biotechniques* 1995;18:62–63.
27. Kissane J, Robins E. The fluorometric measurement of deoxyribonucleic acid in animal tissues with special reference to the central nervous system. *Journal of Biological Chemistry* 1958;233:184–188.
28. Schmidt D, Ernst JD. A fluorometric assay for the quantification of RNA in solution with nanogram sensitivity. *Analytical Biochemistry* 1995;232:144–146.
29. Aplenc R, Orudjev E, Swoyer J, Manke B, Rebbeck T. Differential bone marrow aspirate DNA yields from commercial extraction kits. *Leukemia* 2002;16:1865–1866.
30. Dani S, Gomes-Ruiz AC, Dani MAC. Evaluation of a method for high yield purification of largely intact mitochondrial DNA from human placentae. *Genetic and Molecular Research* 2003;2:178–184.
31. Leal-Klevezas D, Martínez-Vázquez IO, Cuevas-Hernández B, Martínez-Soriano JP. Antifreeze solution improves DNA recovery by preserving the integrity of pathogen-infected blood and other tissues. *Clinical and Diagnostic Laboratory Immunology* 2000;7:945–946.
32. O'Rourke D, Hayes MG, Carlyle SW. Ancient DNA studies in physical anthropology. *Annual Review of Anthropology* 2000;29:217–242.
33. Shia S-R, Cotea RJ, Wub L, Liua C, Datara R, Shia Y, Liua D, Lima H, Taylora CR. DNA extraction from archival formalin-fixed, paraffin-embedded tissue sections based on the antigen retrieval principle: Heating under the influence of pH. *Journal of Histochemistry and Cytochemistry* 2002;50:1005–1011.
34. Cao W, Hashibe M, Rao J-Y, Morgenstern H, Zhang Z-F. Comparison of methods for DNA extraction from paraffin-embedded tissues and buccal cells. *Cancer Detection and Prevention* 2003;27:397–404.
35. McOrist A, Jackson M, Bird AR. A comparison of five methods of extraction of bacterial DNA from human faecal samples. *Journal of Microbiological Methods* 2002;50:131–139.
36. Barbaric D, Dalla-Pozza L, Byrne JA. A reliable method for total RNA extraction from frozen human bone marrow samples taken at diagnosis of acute leukaemia. *Journal of Clinical Pathology* 2002;55:865–867.
37. Byers R, Roebuck J, Sakhinia E, Hoyland J. PolyA PCR amplification of cDNA from RNA extracted from formalin-fixed paraffin-embedded tissue. *Diagnostic Molecular Pathology* 2004;13:144–150.
38. Medeiros M, Sharma VK, Ding R, Yamaji K, Li B, Muthukumar T, Valderde-Rosas S, Hernandez AM, Munoz R, Suthanthiran M. Optimization of RNA yield, purity and mRNA copy number by treatment of urine cell pellets with RNAlater. *Journal of Immunological Methods* 2003;279:135–142.

Chapter 4

Resolution and Detection of Nucleic Acids

Outline

- **ELECTROPHORESIS OF NUCLEIC ACIDS**
- **GEL SYSTEMS**
  - *Agarose Gels*
    - Pulsed-Field Gel Electrophoresis
  - *Polyacrylamide Gels*
- **CAPILLARY ELECTROPHORESIS**
- **BUFFER SYSTEMS**
  - *Buffer Additives*
- **ELECTROPHORESIS EQUIPMENT**
  - *Gel Loading*
- **DETECTION SYSTEMS**
  - *Fluorescent Dyes*
    - Intercalating Agents
    - Minor Groove–Binding Dyes
  - *Silver Stain*

Objectives

4.1 Explain the principle and performance of electrophoresis as it applies to nucleic acids.

4.2 Compare and contrast the agarose and polyacrylamide gel polymers commonly used to resolve nucleic acids, and state the utility of each polymer.

4.3 Explain the principle and performance of capillary electrophoresis as applied to nucleic acid separation.

4.4 Give an overview of buffers and buffer additives used in electrophoretic separation, including the constituents, purpose, and importance.

4.5 Describe the general types of equipment used for electrophoresis and how samples are introduced for electrophoretic separation.

4.6 Compare and contrast pulsed-field gel electrophoresis and regular electrophoresis techniques with regard to method and applications.

4.7 Describe detection systems used in nucleic acid applications.

Resolution and detection of nucleic acids are done in several ways. Gel and capillary electrophoresis are the most practical and frequently used methods. DNA can also be spotted and detected using specific hybridization probes, as is described in Chapter 9, "DNA Sequencing."

ELECTROPHORESIS OF NUCLEIC ACIDS

Electrophoresis is the movement of molecules by an electric current. This can occur in air or solution, or in a matrix to limit migration and contain the migrating material. Electrophoresis is routinely applied to the analysis of proteins and nucleic acids. Each phosphate group on a nucleic acid polymer is ionized, making the molecule negatively charged. Under an electric current, DNA and RNA will migrate toward the positive pole **(anode).** In a matrix of **agarose** or **polyacrylamide,** migration under the pull of the current is impeded, depending on the size of the molecules and the spaces in the **gel** matrix. Because each nucleotide has one negative charge, the charge-to-mass ratio of molecules of different sizes will remain constant. DNA and RNA will therefore migrate at speeds inversely related to the size or length of the polymer. Electrophoresis is performed in tubes, slab gels, or capillaries. Slab gel electrophoresis has either a horizontal or vertical format (Fig. 4.1).

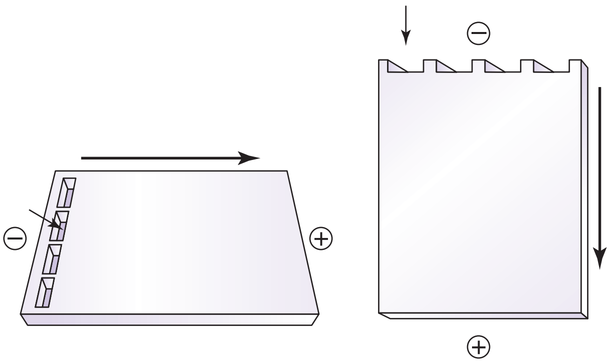


FIGURE 4.1 Horizontal *(left)* and vertical *(right)* gel electrophoresis. In both formats, the sample is introduced into the gel at the cathode end (small arrows) and migrates with the current toward the anode.

Advanced Concepts

Double-stranded DNA and RNA are analyzed by native gel electrophoresis. The relationship between size and speed of migration can be improved by separating single-stranded nucleic acids; however, both DNA and RNA favor the double-stranded state. When complementary strands are not available, single strands will fold and form hydrogen bonds, forming hairpin-like structures, and imperfectly complementary strands will form **heteroduplexes.** This intrastrand or imperfect interstrand hydrogen bonding will impair migration. Unpaired, or denatured, DNA and RNA must therefore be analyzed in conditions that prevent the hydrogen bonding between complementary sequences. These conditions are maintained through a combination of formamide mixed with the sample, urea mixed with the gel, and/or heat.

Historical Highlights

Arne Tiselius developed an electrophoresis apparatus while studying resolution diffusion and adsorption of proteins in naturally occurring porous silicate particles. This "moving-boundary" method required equipment over 5 m long. By the early 1930s, his method was further refined.[1] In the early 1950s, granular starch was introduced as a matrix.[2] Oliver Smithies later found that if he heated about 15 grams of starch in 100 mL of his electrophoretic medium and allowed it to cool, he could make a gel matrix with resolving properties similar to a paper electrophoresis method he had previously devised. He first used this gel to separate serum proteins.[3] In 1956, Smithies and Poulik described the resolution of 20 serum protein components using a two-dimensional electrophoresis system with paper in one dimension and starch in the other.[4]

GEL SYSTEMS

Gel matrices provide resistance to the movement of molecules under the force of an electric current. They prevent diffusion and reduce convection currents so that the separated molecules form a defined group, or **“band.”** The gel can then serve as a support medium for analysis of the separated components. The best matrix would be unaffected by electrophoresis, simple to prepare, and amenable to modification. Agarose and polyacrylamide are polymers that meet these criteria.

Agarose Gels

Agarose is a polysaccharide polymer extracted from seaweed. It is a component of agar used in bacterial culture dishes, which is comprised of agaropectin and agarose. Agarose is a linear polymer of agarobiose, which consists of 1,3-linked-β-d-galactopyranose and 1,4-linked 3, 6-anhydro-α-l-galactopyranose (Fig. 4.2).

Hydrated agarose gels in various concentrations, buffers, and sizes can be purchased ready for use. Alternatively, agarose is purchased and stored in the laboratory in powdered form. For use, powdered agarose is suspended in buffer, heated, and poured into a mold. The concentration of the agarose dictates the size of the spaces in the gel (100 to 300 nm). The size of DNA to be resolved is considered in determining the appropriate agarose concentration to use (Table 4.1). Small pieces of DNA (50 to 500 base pairs [bp]) are resolved on more concentrated agarose gels, for example, 2% to 3% (Fig. 4.3). Larger fragments of DNA (2,000 to 50,000) are best resolved in lower agarose concentrations, for example, 0.5% to 1%. Agarose concentrations above 5% and below 0.5% are not practical. High-concentration agarose will impede migration, whereas very low concentrations produce a weak gel that is easily broken. The gel strength of any concentration of agarose will also decrease over time and with exposure to chaotropic agents such as urea.

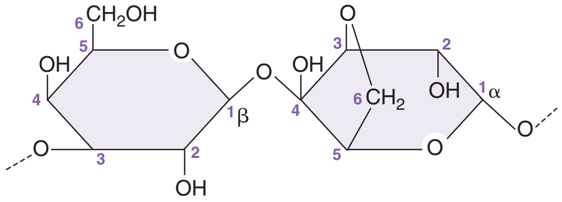


FIGURE 4.2 Agarobiose is the repeating unit of the agarose polymer.

TABLE 4.1 Choice of Agarose Concentration for DNA Gels*[14]

Agarose Concentration (%)	Separation Range (size in bp)
0.3	5,000–60,000
0.6	1,000–20,000
0.8	800–10,000
1.0	400–8,000
1.2	300–7,000
1.5	200–4,000
2.0	100–3,000

*The table shows the range of separation for linear double-stranded DNA molecules in TAE agarose gels with regular power sources. Note that these values may be affected if another running buffer is used and if the voltage is over 5 V/cm.

Advanced Concepts

The physical characteristics of the agarose gel can be modified by altering its polymer length and helical parameters. Several types of agarose are thus available for specific applications. The resolving properties differ in these preparations as well as the gelling properties. Low-melting-point agarose is often used for re-isolating resolved fragments from the gel. Other agarose types give better resolution of larger or smaller fragments. Agarose preparations are sufficiently pure to avoid problems such as **electroendosmosis,** a solvent flow toward one of the **electrodes,** usually the **cathode** (negative), in opposition to the DNA or RNA migration, which slows and distorts the migration of the samples, reducing resolution and smearing the bands.

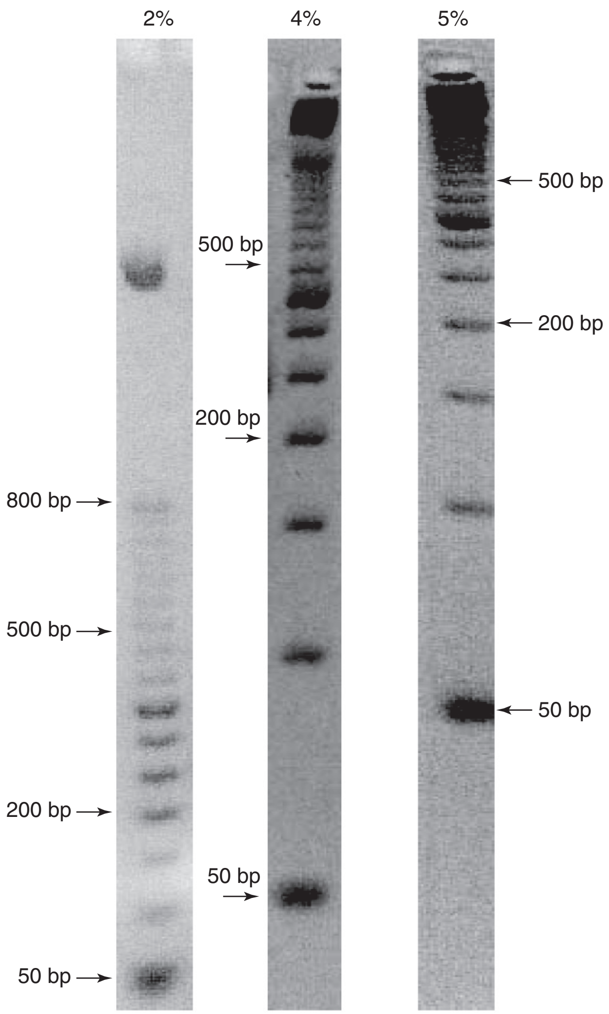


FIGURE 4.3 Resolution of double-stranded DNA fragments on 2%, 4%, and 5% agarose. As the gel concentration increases, the movement of particles (DNA fragments) slows.

Pulsed-Field Gel Electrophoresis

Very large pieces (50,000 to 250,000+ bp) of DNA cannot be resolved efficiently by simple agarose electrophoresis. Even in the lowest concentrations of agarose, megabase fragments are too severely impeded for correct resolution (referred to as limiting mobility). Limiting mobility is reached when a DNA molecule is of such a size that it can move only lengthwise through successive pores of the gel.

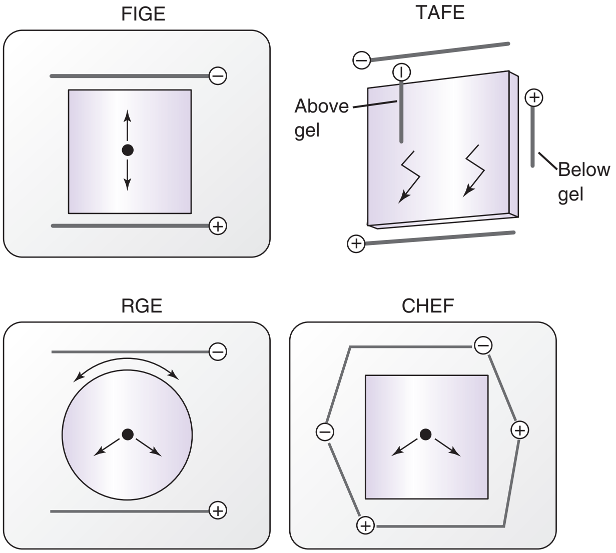


FIGURE 4.4 Field-inversion gel electrophoresis (FIGE), transverse alternating-field electrophoresis (TAFE), rotating gel electrophoresis (RGE), and contour-clamped homogeneous electric field (CHEF) are all examples of pulsed-field gel configurations. Arrows indicate the migration paths of the DNA.

For these very large DNA molecules, pulses of current applied to the gel in alternating dimensions enhance migration. This process is called **pulsed-field gel electrophoresis (PFGE)** (Fig. 4.4). The simplest approach to this method is **field-inversion gel electrophoresis (FIGE)**.[5] FIGE works by alternating the positive and negative electrodes during electrophoresis. In this type of separation, the DNA goes periodically forward and backward. FIGE requires temperature control and a switching mechanism. Field inversion or reorientation to move very large molecules is performed in other configurations. Contour-clamped homogeneous electric field,[6] transverse alternating-field electrophoresis,[7] and rotating gel electrophoresis[8,9] are examples of transverse angle-reorientation electrophoresis methods.

Field-inversion/reorientation systems require gel box, electrode, and gel configurations as well as appropriate electronic control to accommodate switching the electric fields during electrophoresis. Using various forms of PFGE, the large fragments are resolved, not only by sifting through the spaces in the polymer but also by their reorientation. PFGE methods require long separation times because the electrophoresis requires low voltage and the time necessary for the DNA molecules to realign themselves to move in a second dimension, usually an angle of 120° (180° for FIGE) from the original direction of migration.

Advanced Concepts

FIGE is a special modification of PFGE in which the alternating currents are aligned 180° with respect to each other. The current pulses must be applied at different strengths and/or durations so that the DNA will make net progress in one dimension. The parameters for this type of separation must be matched to the DNA being separated so that both large and small fragments have time to reorient properly. For example, if timing is not sufficient for reorientation of the large fragments, small fragments will preferentially reorient and move backward and gradually lose distance with respect to the large molecules, which will continue forward progress on the next pulse cycle. Unlike PFGE that requires special equipment, FIGE can be performed in a regular gel apparatus; however, its upper resolution limit is 2 megabases compared with 5 megabases for PFGE.

The large molecules resolved by these methods must be protected from breakage and shearing. Therefore, specimens for PFGE analysis are immobilized in an agarose plug before cell lysis. Further purification and enzymatic digestion of the DNA are also performed while the DNA is immobilized in the agarose plug. After treatment, the plug is inserted directly into a space or well in the agarose gel for electrophoresis. PFGE instruments will then apply current in alternating directions at specific times (called the switch interval) that are set by the operator. These parameters are based on the general size of the fragments to be analyzed; that is, a larger fragment will require a longer switch interval. PFGE is a slow migration method. Sample runs will take 24 hours or more.

Alternating-field electrophoresis is used for applications that require the resolution of chromosome-sized fragments of DNA, such as in bacterial typing for epidemiological purposes. Enzymatic digestion of genomic DNA will yield a set of fragments that produce a band pattern specific to each type of organism. By comparing band patterns, the similarity of organisms isolated from various sources can be assessed. This information is especially useful in determining the epidemiology of infectious diseases, for example, identifying whether two biochemically identical isolates have a common source.[10]

Polyacrylamide Gels

Very small DNA fragments, single-stranded DNA, RNA, and proteins are best resolved on polyacrylamide gels in polyacrylamide gel electrophoresis (PAGE). Acrylamide, in combination with the cross-linker methylene bisacrylamide (Fig. 4.5), polymerizes into a matrix that has consistent resolution characteristics (Fig. 4.6).

Unlike agarose, which is a natural polymer from living organisms, polyacrylamide is a synthetic material. This allows precise control of the polymer properties and higher resolution than can be achieved with agarose.

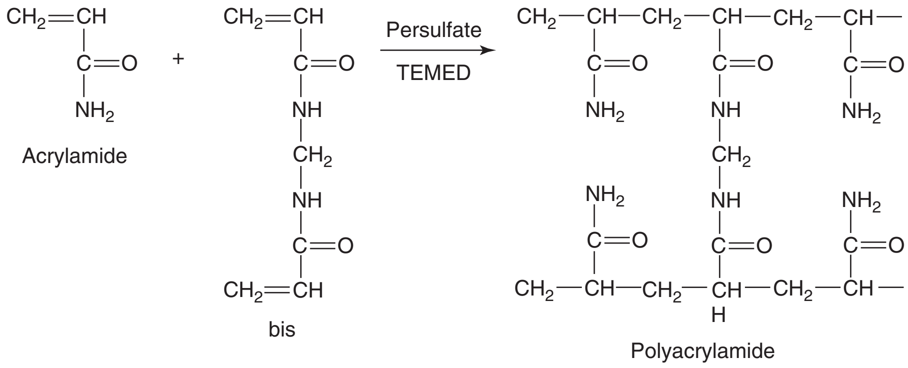


FIGURE 4.5 The repeating unit of polyacrylamide is acrylamide; bis introduces branches into the polymer. In addition to concentration, the acrylamide:bis ratio will determine the degree of branching and the sieving properties of the polymer.

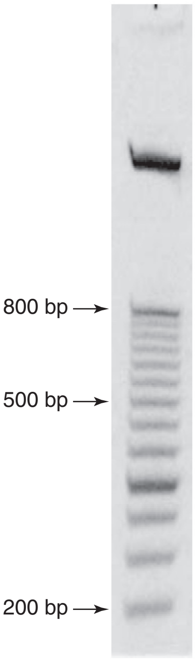


FIGURE 4.6 Resolution of double-stranded DNA fragments on a 5%, 19:1 acrylamide:bis gel.

With single-base resolution, polyacrylamide gels are used for nucleic acid sequencing, mutation analyses, nuclease protection assays, and other applications requiring high resolution of nucleic acids. Acrylamide is supplied to the laboratory in several forms. The powdered form is a dangerous neurotoxin and must be handled with care. Commercial solutions of mixtures of acrylamide and bis-acrylamide are less hazardous and more convenient to use. Preformed gels are the most convenient because the procedure for preparation of acrylamide gels is more difficult than that for agarose gels.

Advanced Concepts

Different cross-linkers affect the physical nature of the acrylamide mesh. Piperazine diacrylate can reduce the background staining that may occur when the gel is stained. *N,N′*-bisacrylylcystamine and *N,N′,z*-diallyltartardiamide enable gels to be solubilized to facilitate extraction of separated products.

The composition of polyacrylamide gels is represented as the total percentage concentration (w/v) of monomer (acrylamide with cross-linker), T, and the percentage of monomer that is cross-linker, C. For example, a 6% 19:1 acrylamide:bis gel has a T value of 6% and a C value of 1/20, or 5%.

Unlike agarose gels that polymerize upon cooling, polyacrylamide gels require a catalyst. The catalyst may be the nucleating agents ammonium persulfate **(APS)** plus *N,N,N′,N′*-tetramethylethylenediamine **(TEMED)** or light activation. APS produces free oxygen radicals in the presence of TEMED to drive the polymerization mechanism. Alternatively, free radicals are generated by a photochemical process using riboflavin plus TEMED. Excess oxygen inhibits the polymerization process. Therefore, de-aeration, or the removal of air, of the gel solution is done before the addition of the nucleating agents.

Polyacrylamide gels for nucleic acid separation can be very thin, for example, 50 µm, making gel preparation difficult. Available systems are designed to facilitate the preparation of single and multiple gels. A more convenient, albeit more expensive, alternative is to use preformed polyacrylamide gels to avoid the hazards of working with acrylamide and the labor time involved in gel preparation. Use of preformed gels must be scheduled, keeping in mind the limited shelf life of the product.

The main advantage of polyacrylamide over agarose is the higher resolution capability of polyacrylamide for small fragments. A variation of 1 bp in a 1-kb molecule (0.1% difference) can be detected in a polyacrylamide gel. Another advantage of polyacrylamide is that there is not as much difference in batches obtained from different sources as there is in agarose. Further, altering T and C in a polyacrylamide gel can change the pore size, and therefore the sieving properties, in a predictable and reproducible manner. Increasing T decreases the pore size proportionally. The minimum pore size (highest resolution for small molecules) occurs at a C value of 5%. Variation of C above or below 5% will increase pore size. Usually, C is set at 3.3% (29:1) for native and 5% (19:1) for standard DNA and RNA gels (see Table 4.2).

CAPILLARY ELECTROPHORESIS

The widest application of capillary electrophoresis has been in the separation of organic chemicals such as

TABLE 4.2 Choice of Polyacrylamide Concentration for DNA Gels*[14]

Acrylamide Concentration (%)	Separation Range (size in bp)
3.5	100–1,000
5.0	80–500
8.0	60–400
12.0	40–200
20.0	10–100

*The indicated figures are referring to gels run in TBE buffer. Voltages over 8 V/cm may affect these values.

pharmaceuticals and carbohydrates. It has also been applied to the separation of inorganic anions and metal ions. For these applications, capillary electrophoresis has the advantages of faster analytical runs and lower cost per run than other separation methods. Capillary electrophoresis is also used for the separation and analysis of nucleic acids.

In this type of electrophoresis, the analyte is resolved in a thin glass (fused silica) capillary that is 30 to 100 cm in length with an internal diameter of 25 to 100 µm. Fused silica is used as the capillary tube because it is the most transparent material allowing for the passage of fluorescent light. The fused silica is covered with a polyimide coating for protection, except for an uncoated window where laser light excites the fluorescent molecules attached to the fragments as they pass a detection device. The capillary has a negative charge along its walls, generated by the dissociation of hydroxyl ions from the molecules of silicone. This establishes an electro-osmotic flow when a current is introduced along the length of the capillary. Under the force of the current, small and negatively charged molecules migrate faster than large and positively charged molecules (Fig. 4.7). Optimal separation requires the use of the proper buffer to ensure that the solute being separated is charged.

Capillary electrophoresis was originally applied to molecules in solution. Separation was based on their size and charge (charge/mass ratio). Negatively charged molecules are completely ionized at high pH, whereas positively charged solutes are completely protonated in low-pH buffers.

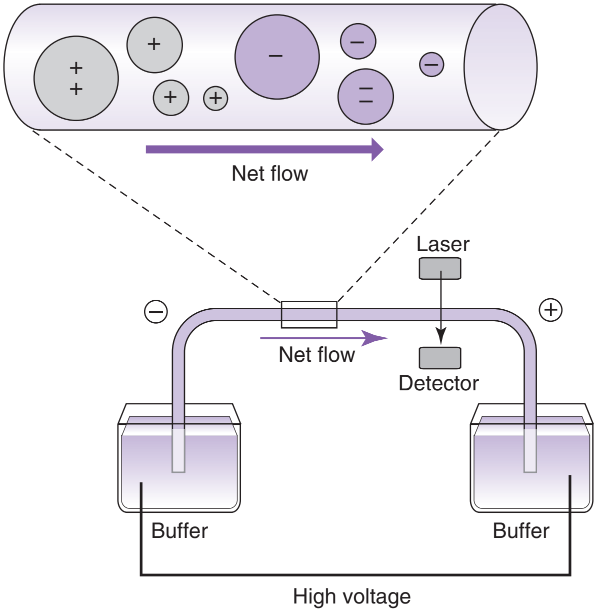


FIGURE 4.7 Capillary electrophoresis separates particles by size (small, fast migration; large, slow migration) and charge (negative, fast migration; positive, slow migration). Because the size and charge of DNA work counter to each other, a polymer (gel) in the capillary will resolve the DNA fragments mostly according to size.

Nucleic acids do not separate well in solution. As the size or length of a nucleic acid increases (slowing migration), so does its negative charge (speeding migration), effectively confounding the charge/mass resolution. Introducing a polymer inside the capillary establishes resolution by impeding nucleic acid migration according to size more than charge. It is important that the nucleic acid be completely denatured (single stranded) so that it will be separated according to its size because the secondary structure will affect the migration speed.

Fluorescent labels are covalently attached to nucleic acids to be separated by capillary electrophoresis. This can be done enzymatically or by polymerase chain reaction priming of DNA synthesis across the region of interest with single-stranded DNA primers, one of which carries a fluorescent molecule at the 5′ end. These processes are described in Chapter 6, “Nucleic Acid Amplification,” and Chapter 9, “DNA Sequencing.” Generally, picogram to nanogram quantities of fluorescently labeled, denatured nucleic acid in buffer containing formamide are introduced to the capillary, which is held at a denaturing temperature, usually 50°C to 60°C.

The sample goes into the capillary through electrokinetic, hydrostatic, or pneumatic injection. **Electrokinetic injection** is used for nucleic acid analysis. In this process, a platinum electrode close to the end of the capillary undergoes a transient high-positive charge to draw the sample to the end of the capillary. When the current is established with a positive charge on the opposite end of the capillary, the nucleic acids migrate into and through the capillary. Capillary electrophoresis is analogous to gel electrophoresis with regard to the electrophoretic parameters for the resolution of nucleic acids. The capillary's small volume, as compared with that of a slab gel, can dissipate heat more efficiently during the electrophoresis process. More efficient heat dissipation allows running the samples at a higher charge per unit area, which means that the samples migrate faster, thereby decreasing the resolution (run) time.

Nucleic acid resolution by capillary electrophoresis is used extensively in forensic applications and parentage testing performed by analyzing short DNA fragments. It has other applications in the clinical laboratory, such as clonality testing, microsatellite instability detection, and bone marrow engraftment analysis. Specially designed software can use differentially labeled molecular-weight markers or allelic markers that, when run through the capillary with the sample, help to identify sample bands.

Compared with traditional slab gel electrophoresis, the capillary system has the advantages of increased sensitivity and immediate detection. With multiple color-detection systems, standards, controls, and test samples are run through the capillary together, thereby eliminating the lane-to-lane variations that can occur across a gel. Although instrumentation for capillary electrophoresis is costly and detection requires fluorescent labeling of samples, labor and runtime are greatly decreased as compared with gel electrophoresis. In addition, analytical software programs automatically analyze the results that are gathered by the detector in the capillary electrophoresis instrument.

BUFFER SYSTEMS

The purpose of a buffer system is to carry the current and protect the samples during electrophoresis. This is accomplished through the electrochemical characteristics of the buffer components.

A buffer is a solution of a weak acid and its **conjugate** base. The pH of a buffered solution remains constant as the buffer molecules take up or release protons given off or absorbed by other solutes. The equilibrium between acid and base in a buffer is expressed as the dissociation constant, K_a:

$$K_a = \frac{[H^+][A^-]}{[HA]}$$

where $[H^+]$, $[A^-]$, and $[HA]$ represent the dissociated proton, conjugate base, and nondissociated acid concentrations, respectively. K_a is most commonly expressed as its negative logarithm, pK_a, such that

$$pK_a = -\log K_a$$

A pK_a of 2 ($K_a = 10^{-2}$) favors the release of protons. A pK_a of 12 ($K_a = 10^{-12}$) favors the association of protons.

A given buffer maintains the pH of a solution near its pKa. The amount the pH of a buffer will differ from the pKa is expressed as the Henderson–Hasselbalch equation:

$$pH = pK_a + \log \frac{[\text{basic form}]}{[\text{acidic form}]}$$

If the acidic and basic forms of the buffer in solution are of equal concentration, $pH = pK_a$. If the acidic form predominates, the pH will be less than the pK_a; if the basic form predominates, the pH will be greater than the pK_a.

The Henderson–Hasselbalch equation predicts that in order to change the pH of a buffered solution by one point, either the acidic or basic form of the buffer must be brought to a concentration of 1/10 that of the other form. Therefore, the addition of acid or base will barely affect the pH of a buffered solution as long as the acidic or basic forms of the buffer are not depleted.

Control of the pH of a gel by the buffer protects sample molecules from damage. Furthermore, the current through the gel is carried by buffer ions, preventing severe fluctuations in the pH of the gel. A buffer concentration must be high enough to provide sufficient acidic and basic forms to buffer its solution. Raising the buffer concentration, however, also increases the conductivity of the electrophoresis system, generating more heat at a given voltage. This can lessen gel stability and increase sample denaturation. High buffer concentrations must therefore be offset by low voltage.

In order for nucleic acids to migrate properly, the gel must be immersed in a buffer that conducts the electric current efficiently in relation to the buffering capacity. Ions with high-charge differences, +2, –2, +3, and so on, move through the gel more quickly; that is, they increase conductivity without increasing buffering capacity. This results in too much current passing through the gel as well as faster depletion of the buffer. Therefore, buffer components such as Tris base [2-amino-2-(hydroxymethyl)-1,3-propanediol] or borate are preferred because they remain partly uncharged at the desired pH and thus maintain constant pH without high conductivity. In addition to pKa, charge, and size, other buffer characteristics that might be taken into account when choosing a buffer include toxicity, interaction with other components, solubility, and ultraviolet (UV) absorption.

Advanced Concepts

The Henderson–Hasselbalch equation also predicts the concentration of the acidic or basic forms at a given pH. It can be used to calculate the state of ionization of a species in solution, that is, the predominance of acidic or basic forms. A buffer should be chosen that has a pK_a within a half point of the desired pH, which is about 8.0 for nucleic acids.

Advanced Concepts

The migration of buffer ions is not restricted by the gel matrix, so the speed of their movement under a current is governed strictly by the size of the ion and its charge (charge/mass ratio). Tris is a relatively large molecule, with a low charge-to-mass ratio, and it moves through the current relatively slowly, even at high concentrations, giving increased buffering capacity.

The Tris buffers, Tris borate EDTA (TBE; 0.089 M Tris-base, 0.089 M boric acid, 0.0020 M EDTA), Tris phosphate EDTA (TPE; 0.089 M Tris-base, 1.3% phosphoric acid, 0.0020 M EDTA), and Tris acetate EDTA (TAE; 0.04 M Tris-base, 0.005 M sodium acetate, 0.002 M EDTA), are most commonly used for electrophoresis of DNA. TBE has a greater buffering capacity than TAE. Although the ion species in TAE are more easily exhausted during extended or high-voltage electrophoresis, DNA will migrate twice as fast in TAE than in TBE in a constant current. When using any buffer, especially TBE and TPE, the gel can overheat when run at high voltage in a closed container. Finally, stock solutions of TBE are prone to precipitation. This can result in differences in concentration between the buffer in the gel and the running buffer. Such a gradient will cause localized distortions in nucleic acid migration patterns, often causing a salt wave that is visible as a sharp horizontal band through the gel.

Buffer Additives

Buffer additives modify sample molecules in ways that affect their migration. Examples of these additives are formamide, urea, and detergents. With regard to nucleic acids, **denaturing agents**, such as formamide or urea, break hydrogen bonds between complementary strands or within the same strand of DNA or RNA. The conformation or solubility of molecules can be standardized by the addition of one or both of these agents. Formamide and heat added to DNA and RNA break and block the hydrogen-bonding sites, hindering complementary sequences from reannealing. As a result, the molecules become long, straight, unpaired chains. Urea and heat in the gel systems maintain this conformation such that intrachain hybridization (folding) of the nucleic acid molecules does not affect migration speeds, and separation occurs strictly according to the size or length of the molecule.

Electrophoresis of RNA requires different conditions imparted by different additives from those that are used with DNA. Because RNA is single stranded and longer molecules tend to fold to optimize internal hydrogen bonding, they must be completely denatured to prevent folding in order to accurately determine the size by migration in a gel system. The secondary structures formed in RNA are strong and more difficult to denature than DNA homologies. Denaturants used for RNA include methylmercuric hydroxide (MMH), which reacts with amino groups on the RNA to prevent **base pairing** between complementary nucleotides and with

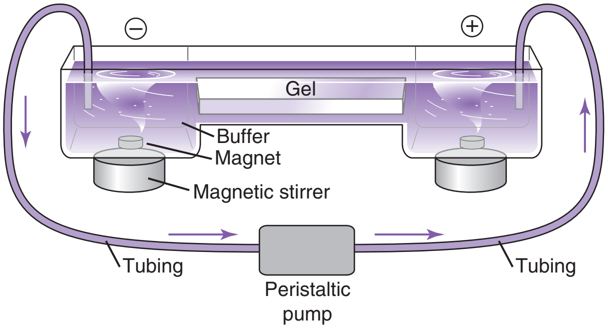


FIGURE 4.8 A peristaltic pump is used to recirculate buffer from the cathode to the anode end while running a denaturing gel.

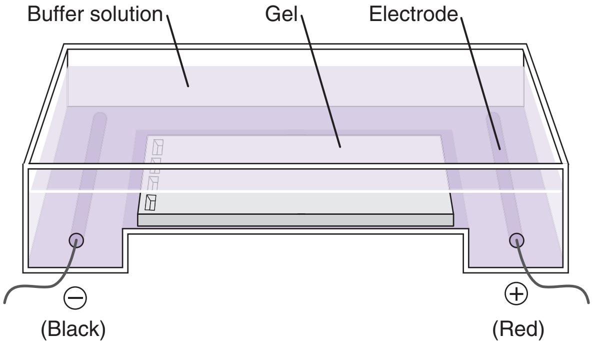


FIGURE 4.9 A typical horizontal submarine gel system. A red connector is attached to the positive outlet on the power supply, and a black connector is attached to the negative port. Nucleic acid will migrate to the positive (red) pole.

aldehydes (e.g., formaldehyde, glyoxal), which also disrupt base pairing. MMH is not used routinely because of its extreme toxicity.

Examples of RNA electrophoresis buffers are 10-mM sodium phosphate, pH 7, and MOPS buffer [20 mM 3-(*N*-morpholino) propanesulfonic acid, pH 7, 8 mM sodium acetate, 1 mM EDTA, pH 8]. The RNA sample is incubated in dimethyl sulfoxide, 1.1 M glyoxal (ethane 1.2 dione) and 0.01 M sodium phosphate, pH 7, to denature the RNA prior to loading the sample on the gel. Due to **pH drift** (decrease of pH at the cathode [–] and an increase of the pH at the anode [+]) during the run, the buffer may have to be recirculated from the anode end of the bath to the cathode end (Fig. 4.8). This is accomplished by using a peristaltic pump or stopping the gel at intervals and transferring the buffer from the cathode to the anode ends.

ELECTROPHORESIS EQUIPMENT

Gel electrophoresis is performed in one of two conformations, horizontal or vertical. In general, agarose gels are run horizontally, and polyacrylamide gels are run vertically. This, however, is not always the case.

Horizontal gels are run in acrylic gel boxes or baths that are divided into two parts, with a platform in the middle on which the gel rests (Fig. 4.9). The electrodes (platinum wires) placed across the box at each end of the bath compartments are connected to a power supply through the walls of the container. The gel is submerged in electrophoresis buffer that fills both compartments and makes a continuous system through which the current flows. The thickness of the gel and the volume of the buffer affect the current, and therefore the migration of the sample, so these parameters are kept constant for consistent results. Because the gel is submerged throughout the loading and electrophoresis process, horizontal gels are sometimes referred to as **submarine gels.**

Once the sample is introduced into the gel, the electrodes are connected to the power source. The power supply will deliver voltage, setting up a current that will run through the gel buffer and the gel, carrying the charged sample through the gel matrix at a speed corresponding to the charge/mass ratio of the sample molecules.

Horizontal agarose gels are cast as square or rectangular slabs of varying size. Commercial gel boxes come with casting trays that mold the gel to the appropriate size for the gel box. The volume of the gel solution will determine the thickness of the gel. Agarose, supplied as a dry powder, is mixed at a certain percentage (w/v) with electrophoresis buffer and heated on a heat block or by microwave to dissolve and melt the agarose. The molten agarose is cooled to between 55°C and 65°C, and the appropriate volume is poured into the casting tray as dictated by the gel box manufacturer or application. A **comb** is then inserted into the top of the gel to create holes, or **wells,** in the gel into which the sample will be loaded. The size of the teeth in the comb will determine the capacity of the well for the sample, and the number of teeth in the comb will determine the number of wells that are available in the gel to receive the samples. The gel is then allowed to cool, during which time it will solidify. After the gel has polymerized, the comb is carefully removed, and the gel is placed into the gel box and submerged in the electrophoresis buffer.

There is a significant shock hazard if contact is made with the gel buffer while the current is on. For safety purposes, gel baths are designed to stop the current when a protective cover is removed from the gel bath. Enclosed cassettes containing gel, buffer, electrodes, and detection do not require submersion in running buffer. The sample is loaded directly into the wells of these systems.

Vertical gels are cast between glass plates that are separated by spacers. The spacers determine the thickness of the gel, ranging from 0.05 to 4 mm. The bottom of the gel is secured by tape or by a gasket in specially designed gel casting trays. After the addition of catalyst and nucleating agents, the liquid acrylamide is poured or forced between the glass plates with a pipet or a syringe. The comb is then placed on the top of the gel. For light-activated polymerization, the gel between the glass plates is exposed to a light source. During this process, it is important to prevent air from getting into the gel or beneath the comb. Bubbles will form discontinuities in the gel, and oxygen will inhibit the polymerization of the acrylamide. The comb is of a thickness equal to that of the spacers so that the gel will be the same thickness throughout.

As with horizontal gels, the number and size of the comb's teeth determine the number of wells in the gel and the sample volume that can be added to each well. Specialized combs, called **sharkstooth combs,** facilitate lane to lane band comparisons (Fig. 4.10). These combs are placed upside down (teeth up, not in contact with the gel) to form a straight surface on the gel during polymerization. After polymerization is complete, the comb is removed and placed tooth-side down on top of the gel for loading. With this configuration, the spaces between the comb teeth form the well walls. The advantage to this arrangement is that the resulting gel lanes are immediately adjacent to one another. When used, standard combs are removed before the gel is loaded, whereas the samples are loaded while the sharkstooth comb is in place. When the standard combs are removed from the gel, care must be taken not to break or displace the "ears" that were formed by the spaces between the teeth in the comb that separate the gel wells.

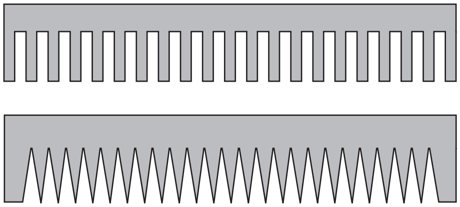


FIGURE 4.10 Combs for polyacrylamide electrophoresis. Regular combs *(top)* have teeth that form the wells in the gel. Sharkstooth combs *(bottom)* are placed onto the polymerized gel, and the sample is loaded between the teeth of the comb.

Vertical gel boxes have separate chambers connected by the gel itself. Electrodes are attached to the upper and lower buffer chambers to set up the current that will run through the gel.

The gel must be in place before filling the upper chamber with buffer. Some systems have a conductive plate attached to the back of the gel to maintain a constant temperature across the gel. Maintaining constant temperature is more of a problem with vertical gels because the outer edges of the gel cool more than the center, slowing migration in the outer lanes compared with lanes in the center of the gel. This is called "gel smiling" because similar-sized bands in the cooler outer lanes will migrate slower than comparable bands in the inside lanes. Vertical gel systems can range from large sequencing gels (35 × 26 cm) to minigels (8 × 10 cm). Some minigel systems accommodate two or more gels at a time (Fig. 4.11). Minisystems are used extensively for analyses that do not require single base-pair resolution. The larger systems have been used for procedures requiring higher resolution but are being replaced by capillary electrophoresis for many applications.

Vertical gels are loaded from the top, below a layer of buffer in the upper chamber. Long, narrow, gel-loading pipette tips that deposit the sample neatly on the floor of the well increase band resolution and sample recovery.

Advanced Concepts

Self-contained agarose gel systems have been developed to facilitate the electrophoresis process. They are manufactured in closed plastic cassettes containing buffer, gel, and stain. These are

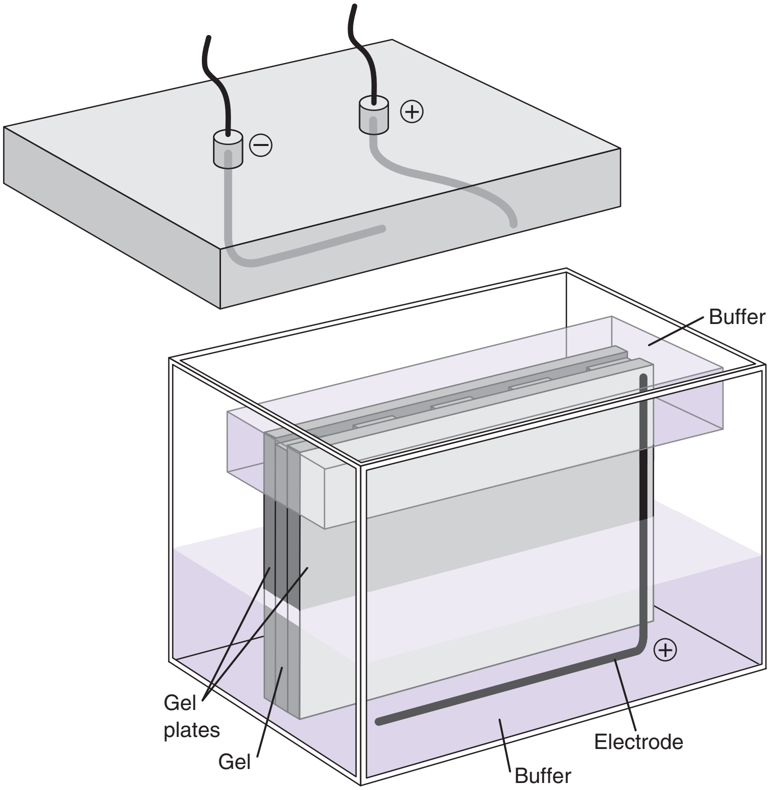


FIGURE 4.11 A typical vertical gel apparatus. Polymerized gels are clamped into the gel insert and placed in the gel bath. The positive electrode will be in contact with the bottom of the gel and the buffer, filling about a third of the gel bath. The negative electrode will be in contact with the top of the gel and a separate buffer compartment in the top of the insert.

convenient for routine use, but they restrict the gel configuration, that is, the number and size of wells, and so forth. Also, the percentage of agarose or acrylamide is limited to what is available from the manufacturer. Furthermore, the separated nucleic acids cannot easily be removed from these closed cassettes, limiting post-electrophoretic procedures.

Polyacrylamide gels can also be cast in tubes for isoelectric focusing or two-dimensional gel electrophoresis. The tubes containing the gels are placed in a chamber separated like those for vertical slab gels. The tubes are held in place by gaskets in the upper chamber. When a tubular gel is placed at the top of a slab gel, the molecules separated in one direction in the tube are then further separated in a second dimension through the slab gel. Two-dimensional gel electrophoresis is mostly applied to protein separations.

Gel Loading

Prior to loading the sample containing isolated nucleic acid onto a gel, **tracking dye** and a density agent are added to the sample. The density agent (Ficoll, sucrose, or glycerol) increases the density of the sample as compared with the electrophoresis buffer. When the sample solution is dispensed into the wells of the gel below the surface of the buffer, it sinks into the well instead of diffusing in the buffer. Although samples are pipetted into the wells as close as possible to the well floor, the density agent mixed with the sample allows loading from

TABLE 4.3 Tracking Dye Comigration*

Gel %	Bromophenol Blue (Nucleotides)	Xylene Cyanol (Nucleotides)
Agarose		
0.5–1.5	300–500	4,000–5,000
2.0–3.0	80–120	700–800
4.0–5.0	20–30	100–200
PAGE		
4	95	450
6	60	240
8	45	160
10	35	120
12	20	70
20	12	45

*Migration depends on buffer type (TAE, TBE, or TPE) and the formulation of agarose, acrylamide, and bis.

far enough away so as not to risk damaging the well. If a well is damaged, the tracking dye in the sample will be seen seeping out of the well and under the gel.

The tracking dyes are used to monitor the progress of the electrophoresis run. The dyes migrate at specific speeds in a given gel concentration and usually run ahead of the smallest fragments of DNA (Table 4.3). Bromophenol blue is a tracking dye that is used for many applications, and xylene cyanol green is another of the chromophores used as tracking dyes for both agarose and polyacrylamide gels. Tracking dyes are not associated with the sample DNA, and thus they do not affect the separation. The movement of the tracking dye is monitored, and when the dye approaches the end of the gel, or the desired distance, electrophoresis is terminated. Running a gel too long will result in loss of some or all of the desired bands because they will run off the bottom of the gel into the gel bath. Timers or automatic power cutoffs are used to avoid loss of sample. Some gel systems will automatically stop at the appropriate time for the gel size and current strength.

Advanced Concepts

A type of "bufferless" electrophoresis system supplies buffer in gel form or strips. These are laid next to the preformed gel on a platform that replaces the electrophoresis chamber. These systems can offer the additional advantage of precise temperature control during the run.

Advanced Concepts

Gels in cassette systems and gel strip systems can be loaded without loading buffer because the wells are "dry," precluding the need for density gradients. These systems also have automatic shut-off at the end of the run, so tracking dye is usually not necessary, although some systems have a tracking dye built into the gel and/or buffer.

DETECTION SYSTEMS

The agents used most frequently for visualization of bands after electrophoresis are fluorescent dyes and silver stain. Both of these types of dyes specifically associate with nucleic acid.

Fluorescent Dyes

Intercalating Agents

Intercalating agents intercalate, or stack, between the nitrogen bases in double-stranded nucleic acid. Ethidium bromide (3,8-diamino-5-ethyl-6-phenylphenanthridinium bromide [EtBr]) is one of these agents and was the most widely used dye in early DNA and RNA analyses. Because EtBr is carcinogenic, precautions are required to limit exposure. Under excitation with UV light at 300 nm, EtBr in DNA emits visible light at 590 nm. Therefore, DNA separated in agarose or acrylamide and exposed to EtBr will emit orange light when illuminated at 300 nm. After electrophoresis, the agarose or acrylamide gel is soaked in a solution of

0.1- to 1-mg/mL EtBr in running buffer (TAE, TBE, or TPE) or in TE. Alternatively, dye is added directly to the gel before polymerization or to the running buffer. The latter two measures save time and allow visualization of the DNA during and immediately after the run; however, dye added to the gel may form a bright front across the gel that could mask informative bands. Dye added to the running buffer produces consistent staining, although greater volumes of hazardous waste are generated by this method. To increase safety, noncarcinogenic dyes with similar excitation and emission wavelengths as EtBr have been developed. Alternatively, enclosed gel systems contain EtBr inside a plastic-enclosed gel cassette, limiting exposure and limiting waste. After soaking or running in dye, the DNA illuminated with UV light will appear as orange bands in the gel. The image is captured with appropriate filters by digital transfer to analytical software.

Minor Groove–Binding Dyes

SYBR green (*N′*,*N′*-dimethyl-*N*-[4-[(E)-(3-methyl-1,3-benzothiazol-2-ylidene)methyl]-1-phenylquinolin-1-ium-2-yl]-*N*-propylpropane-1,3-diamine) is one of a set of stains introduced in 1995 as another type of nucleic acid–specific dye system. The related dyes include SYBR green I used for double-stranded DNA staining; SYBR green II, for single-stranded DNA or RNA staining; and SYBR gold, for both DNA and RNA staining. SYBR green I differs from EtBr in that it does not intercalate between bases; it sits in the minor groove of the double helix. SYBR green in association with DNA or RNA also emits light in the orange range (522 nm). In agarose gel electrophoresis, SYBR staining is 25 to 100 times more sensitive than EtBr (detection level: 60 pg of double-stranded DNA versus 5 ng for EtBr). This is due, in part, to background fluorescence from EtBr in agarose. A 1× dilution of the manufacturer's 10,000× stock solution of SYBR green I in TAE, TBE, or TE is used in the methods described for EtBr. SYBR green can also be added directly to the DNA sample before electrophoresis. DNA prestaining decreases the amount of dye required for DNA visualization but lowers the sensitivity of detection and may, at higher DNA concentrations, interfere with DNA migration through the gel.[11] Because SYBR green is not an intercalating agent, it is not as mutagenic and is therefore safer to use.[12]

SYBR gold stain is a cyanine dye of proprietary structure. Its fluorescence increases more than 1,000-fold upon binding to double- or single-stranded DNA or to RNA. Like SYBR green, SYBR gold is excited by UV light (300-nm wavelength). SYBR gold emits light at 537 nm.

Although SYBR dyes have some advantages over EtBr, many laboratories continue to use the EtBr due to the requirement for special optical filters for detection of SYBR green emission at 520- to 550-nm wavelengths and SYBR gold at 537 nm. Scanning and photographic equipment optimized for EtBr would have to be modified for optimal detection of the SYBR stains. New instrumentation with more flexible detection systems allows utilization of the SYBR and other fluorescent stains. SYBR green is the preferred dye for real-time PCR methods.

Silver Stain

Another sensitive staining system originally developed for protein visualization is silver stain. After electrophoresis, the sample is fixed with methanol and acetic acid. The gel is then impregnated with ammoniacal silver (silver diamine) solutions or silver nitrate in a weakly acid solution.[13] Interaction of silver ions with acidic or nucleophilic groups on the target results in crystallization or deposition of metallic silver under optimal pH conditions. The insoluble black silver salt precipitates upon introduction of formaldehyde in a weak acid solution, or alkaline solution for silver nitrate. Of the two procedures, silver diamine is best for thick gels, whereas silver nitrate is considered to be more stable.[14] Optimization of silver staining has continued to simplify and increase its efficiency. One method reports sensitivity of detection as low as 14.6 pg in 6 to 7 minutes using only three reagents (silver nitrate, sodium hydroxide, and formaldehyde).[15]

Silver staining avoids the hazards of the intercalators, but silver nitrate is also a biohazard and must be handled accordingly. In addition, silver staining is more complicated than simple fluorescent stains. Color development must be carefully watched in order to stop the reaction once the optimal signal is reached. Overdevelopment of the color reaction will result in high backgrounds and masking of results. The increased sensitivity of this staining procedure, however, makes up for its limitations. It is especially useful for protein analysis and for detection of limiting amounts of product.

STUDY QUESTIONS

1. You wish to perform an electrophoretic resolution of your restriction enzyme–digested DNA. The sizes of the expected fragments range from 100 to 500 bp. You discover two agarose gels polymerizing on the bench. One is 0.5% agarose; the other is 2% agarose. Which one might you use to resolve your fragments?

2. After completion of the electrophoresis of DNA fragments along with the proper molecular-weight standard on an agarose gel, suppose the outcomes in (a) and (b) were observed. What might be the explanations for each outcome?
    a. The gel is blank (no bands, no molecular-weight standard).
    b. Only the molecular weight standard is visible.

3. How does PFGE separate larger fragments more efficiently than standard electrophoresis?

4. A 6% solution of 19:1 acrylamide:bis-acrylamide is mixed, de-aerated, and poured between glass plates for gel formation. After an hour, the solution is still liquid. What might be one explanation for the gel not polymerizing?

5. A gel separation of RNA yields aberrantly migrating bands and smears. Suggest two possible explanations for this observation.

6. Why does DNA not resolve well in solution (without a gel matrix)?

7. Why is SYBR green less toxic than EtBr?

8. What are the general components of loading buffer used for introducing DNA samples to submarine gels?

9. Name two dyes that are used to monitor the migration of nucleic acid during electrophoresis.

10. When a DNA fragment is resolved by slab gel electrophoresis, a single sharp band is obtained. What would the equivalent observation be if this fragment had been fluorescently labeled and resolved by capillary electrophoresis?

References

1. Tiselius A. A new apparatus for electrophoretic analysis of colloidal mixtures. *Transactions of the Faraday Society* 1937;33:524.
2. Kunkel H, Slater RJ. Zone electrophoresis in a starch supporting medium. *Proceedings of the Society of Experimental Biology and Medicine* 1952;80:42–44.
3. Smithies O, Walker NF. Genetic control of some serum proteins in normal humans. *Nature* 1955;176:1265–1266.
4. Smithies O, Poulik MD. Two-dimensional electrophoresis of serum proteins. *Nature* 1956;177:1033.
5. Carle G, Frank M, Olson MV. Electrophoretic separation of large DNA molecules by periodic inversion of the electric field. *Science* 1986;232:65–68.
6. Chu G, Vollrath D, Davis RW. Separation of large DNA molecules by contour-clamped homogeneous electric fields. *Science* 1986;234:1582–1585.
7. Gardiner K, Laas W, Patterson DS. Fractionation of large mammalian DNA restriction fragments using vertical pulsed-field gradient gel electrophoresis. *Somatic Cell and Molecular Genetics* 1986;12:185–195.
8. Southern E, Anand R, Brown WRA, Fletcher DS. A model for the separation of large DNA molecules by crossed field gel electrophoresis. *Nucleic Acids Research* 1987;15:5925–5943.
9. Gemmill R. Pulsed field gel electrophoresis. In *Advances of Electrophoresis*. Edited by Chrambach A, Dunn MJ, Radola, BJ. Weinheim, Germany: VCH, 1991. pp. 1–48.
10. Donnenberg M, Narayanan S. How to diagnose a foodborne illness. *Infectious Disease Clinics of North America* 2013;27:3.
11. Miller S, Taillon-Miller P, Kwok P. Cost-effective staining of DNA with SYBR green in preparative agarose gel electrophoresis. *BioTechniques* 1999;27:34–36.
12. Singer V, Lawlor, TE, Yue, S. Comparison of SYBR green I nucleic acid gel stain mutagenicity and ethidium bromide mutagenicity in the *Salmonella*/mammalian microsome reverse mutation assay (Ames test). *Mutation Research* 1999;439:37–47.
13. Rabilloud T. A comparison between low background silver diamine and silver nitrate protein stains. *Electrophoresis* 1992;13:429–439.
14. Merrill C. Gel-staining techniques. *Methods in Enzymology* 1990;182:477–488.
15. Liu W, Li R, Ayalew H, Xia Y, Bai G, Yan G, Siddique KH, Guo P. Development of a simple and effective silver staining protocol for detection of DNA fragments. *Electrophoresis* 2017;38(8):1175–1178.

Chapter 5

Analysis and Characterization of Nucleic Acids and Proteins

Outline

RESTRICTION ENZYME MAPPING OF DNA
CRISPR ENZYME SYSTEMS
HYBRIDIZATION TECHNOLOGIES
  Southern Blots
    Restriction Enzyme Cutting and Resolution
    Preparation of Resolved DNA for Blotting (Transfer)
    Blotting (Transfer)
PROBE HYBRIDIZATION
  Northern Blots
  Western Blots
PROBES
  DNA Probes
  RNA Probes
  Other Nucleic Acid Probe Types
  Protein Probes
  Probe Labeling
  Nucleic Acid Probe Design
HYBRIDIZATION CONDITIONS, STRINGENCY
DETECTION SYSTEMS
INTERPRETATION OF RESULTS
ARRAY-BASED HYBRIDIZATION
  Dot/Slot Blots
  Genomic Array Technology
    Macroarrays
    Microarrays
    Bead Array Technology
SOLUTION HYBRIDIZATION

Objectives

5.1 Describe how restriction enzyme sites are mapped on DNA.
5.2 Construct a restriction enzyme map of a DNA plasmid or fragment.
5.3 Diagram the Southern blot procedure.
5.4 Explain depurination and denaturation of resolved DNA.
5.5 Describe the procedure involved in blotting (transfer) DNA from a gel to a membrane.
5.6 Discuss the purpose and structure of probes that are used for blotting procedures.
5.7 Define *hybridization*, *stringency*, and *melting temperature*.
5.8 Calculate the melting temperature of a given sequence of double-stranded DNA.

5.9 Compare and contrast radioactive and nonradioactive DNA detection methods.
5.10 Compare and contrast dot and slot blotting methods.
5.11 Describe microarray methodology.
5.12 Discuss solution hybridization.

There is a growing assortment of methods for analysis and characterization of nucleic acids. Advances in DNA sequencing technology have made analysis at single-nucleotide resolution relatively straightforward. Sequencing technology is described in Chapter 9, "DNA Sequencing." Even so, some tests require detection of known sequence variations at a single codon or even a single nucleotide, for which the effort and expense of sequencing are less efficient. In these cases, methods derived even before the development of current sequencing technology are most appropriate.

RESTRICTION ENZYME MAPPING OF DNA

Clinical and forensic analyses require characterization of specific genes or genomic regions at the molecular level. Because of their sequence-specific activity, restriction endonucleases provide a convenient tool for molecular characterization of DNA.

Restriction enzymes commonly used in the laboratory (type II restriction enzymes) have 4 to 6 base-pair (bp) recognition sites, or binding/cutting sites, on the DNA. Any 4- to 6-bp nucleotide sequence occurs at random in a sufficiently long stretch of DNA. Therefore, restriction sites will occur in all DNA molecules of sufficient length. The location of restriction sites will differ among DNA molecules with different sequences. Restriction site mapping (i.e., determining where in the DNA sequence a particular restriction enzyme recognition site is located) was developed using small circular bacterial plasmids. The resultant maps were used to identify and characterize naturally occurring plasmids and to engineer the construction of recombinant plasmids.

To make a **restriction map,** DNA is exposed to several restriction enzymes separately and then in different combinations. Take, for example, a linear fragment of DNA digested with the enzyme *Pst*I. After incubation with the enzyme, the resulting fragments are separated by gel electrophoresis. The gel image reveals four fragments, labeled A, B, C, and D, produced by *Pst*I (Fig. 5.1). From the number of fragments, one can deduce the number of *Pst*I sites: three. The sizes of the fragments, as determined by comparison with known molecular-weight standards, indicate the distance between *Pst*I sites or from one *Pst*I site to the end of the fragment. Although *Pst*I analysis of this fragment yields a characteristic four-band restriction pattern, it does not indicate the order of the four restriction products in the original fragment.

To begin to determine the order of the restriction fragments, another enzyme is used, for example, *Bam*HI. Cutting the same fragment with *Bam*HI yields two pieces, indicating one *Bam*HI site in this linear fragment (see Fig. 5.1). In this figure, observe that one restriction product (F) is much larger than the other (E). This means that the *Bam*HI site is close to one end of the fragment. When the fragment is cut simultaneously with *Pst*I and *Bam*HI, five products are produced, with *Pst*I product A cut into two pieces by *Bam*HI. Because the *Bam*H1 site is known to be close to one end of the fragment, *Pst*I fragment A is on one end of the DNA fragment. By measuring the number and length of products produced by other enzymes, the restriction sites can be placed in linear order along the DNA sequence. Figure 5.2 shows two possible maps based on the results of cutting the fragment with *Pst*I and *Bam*HI. With adequate enzymes and enzyme combinations, a detailed map of this fragment is generated.

Mapping of a circular plasmid is slightly different because there are no free ends (Fig. 5.3). The example shown in the figure is a 4-kb-pair circular plasmid with one *Bam*HI site and two *Xho*I sites. Cutting the plasmid with *Bam*HI will yield one fragment. The size of the fragment is the size of the plasmid, 4 kb in this example. Two fragments released by *Xho*I indicate that there are two *Xho*I sites in the plasmid and that these sites are 1.2 and 2.8 kb pairs away from each other. As with linear mapping, cutting the plasmid with *Xho*I and *Bam*HI at the same time will start to order the sites with respect to one another on the plasmid. One possible arrangement is shown in Figure 5.3. As more enzymes are used, the map becomes more detailed.

Under the proper reaction conditions, restriction enzymes are highly specific for their recognition and

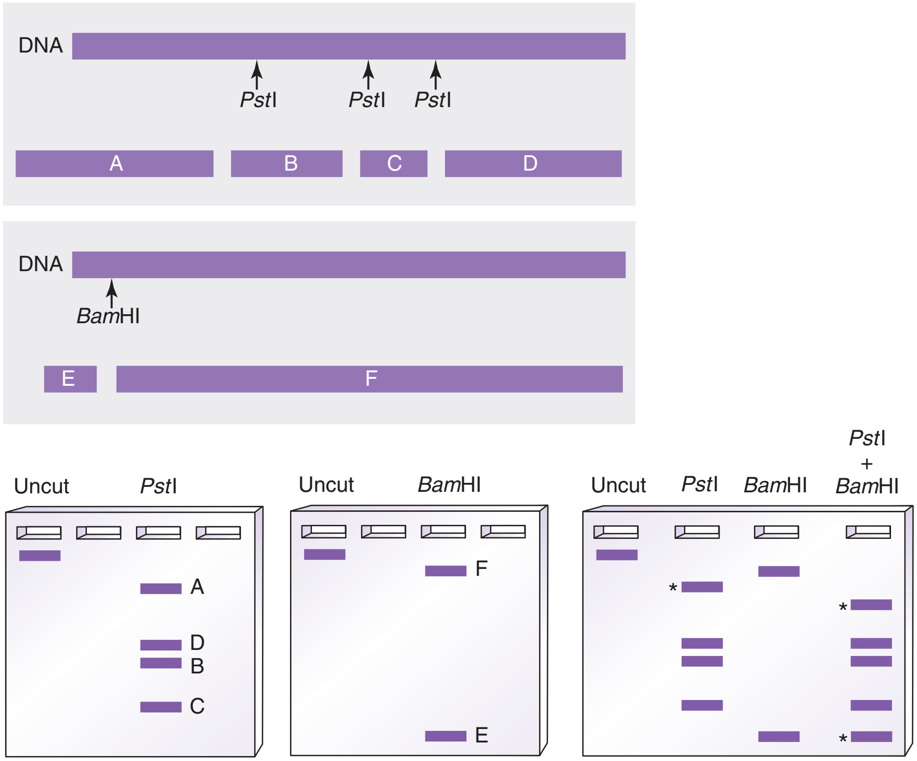


FIGURE 5.1 Restriction mapping of a linear DNA fragment (*top,* green bar). The fragment is first cut with the enzyme *Pst*I. Four fragments result, as determined by agarose gel electrophoresis, indicating that there are three *Pst*I sites in the linear fragment. The size of the pieces indicates the distance between the restriction sites. A second cut with *Bam*HI *(bottom)* yields two fragments, indicating one site. Because one *Bam*HI fragment (E) is very small, the *Bam*HI site must be near one end of the fragment. Cutting with both enzymes indicates that the *Bam*HI site is in the *Pst*I fragment A.

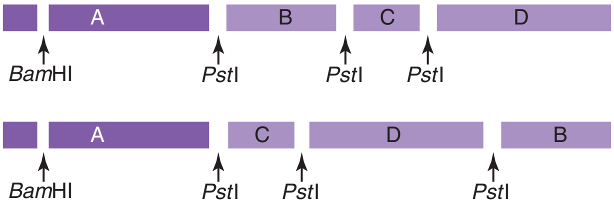


FIGURE 5.2 Two possible maps inferred from the observations described in Figure 5.1. The *Bam*HI site positions fragment A at one end (or the other) of the map. Determination of the correct map requires information from additional enzyme cuts.

cutting properties. Under nonstandard conditions, however, some restriction enzymes will bind to and cut sequences other than the expected recognition sequence. This altered specificity is called star activity. The propensity for star activity varies among enzymes.[1] Thus, the nature and degree of star activity depend on the enzyme and the reaction conditions. Reaction conditions that induce star activity include suboptimal buffer, contamination with organic solvents (e.g., ethanol), high concentrations of glycerol, prolonged reaction time, high concentration of enzymes, pH greater than 8.0, and divalent cation imbalance.

The pattern of fragments produced by restriction enzyme digestion of a DNA fragment or region can be used to identify that DNA. Because of inherited

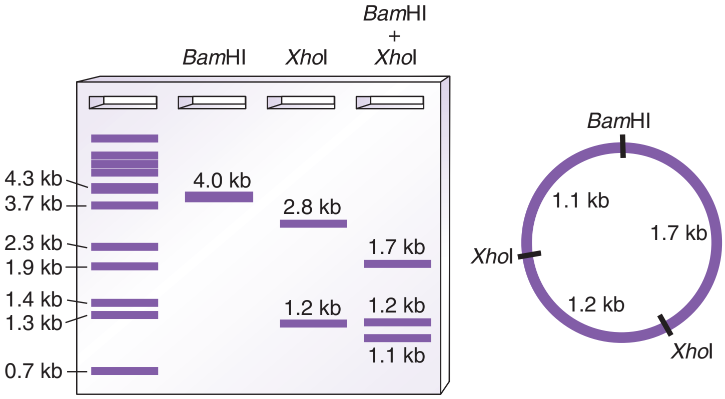


FIGURE 5.3 Restriction mapping of a plasmid. After incubating plasmid DNA with restriction enzymes, agarose gel electrophoresis banding patterns indicate the number of restriction sites and the distance between them.

or somatic differences in the nucleotide sequences in human DNA, the number and location of restriction sites for a given restriction enzyme are not the same in all individuals. The resulting differences in the size or number of restriction fragments are **restriction fragment length polymorphisms (RFLPs).** In addition to their use in epidemiological studies to identify microorganisms and plasmids, RFLPs were the basis of the first molecular-based human identification and mapping methods. RFLPs are also used for the clinical analysis of structural changes in chromosomes associated with disease (translocations, deletions, insertions).

CRISPR ENZYME SYSTEMS

DNA cut sites of restriction enzymes used for DNA analysis are limited to sequences recognized by the enzyme proteins. Another type of restriction system found in archaea, gram-negative bacteria, and gram-positive bacteria guides a common enzyme to specific sites determined by RNA components. This flexible system has proven to be useful for manipulation of both DNA sequence and RNA expression.

Clustered regularly interspaced short palindromic repeats (CRISPRs) are classes of repeated DNA sequences found in prokaryotic and archaebacterial genomes. They are repeated sequences interrupted by spacer sequences matching the genome regions of plasmids or bacteriophages that had previously infected the bacterium. DNA from new invaders is incorporated into the CRISPR locus within a series of short (~20 bp) repeats. These spacer sequences serve as adaptive immunity with memory of the invading DNA. The locus also encodes an endonuclease, CRISPR-associated protein (Cas).

To fend off an invader, short RNA sequences transcribed from the CRISPR spacer regions (CRISPR RNA or **crRNA**) are processed from larger pre-crRNA transcripts. The crRNA then forms a complex with a trans-activating CRISPR RNA **(tracrRNA)** and Cas enzyme (Fig. 5.4). This complex, led to its target by the crRNA homology where it binds and cuts the invading DNA. Cas9 requires a specific protospacer adjacent motif (PAM) to cut the DNA. The protospacer is the part of the crRNA sequence that is complementary to the target sequences incorporated into the bacterial genome. The PAM varies depending on the bacterial species of the Cas gene. The Cas9 nuclease from *Streptococcus pyogenes* recognizes a PAM sequence of NGG directly 3′ of the target sequence in the target DNA, on the complementary strand. The PAM may facilitate the formation of the RNA:DNA hybrid between the crRNA and the target DNA.

There are three types of CRISPR/Cas systems: type I and II that target double-stranded DNA and type III that targets single-stranded DNA and RNA. The crRNAs from each CRISPR locus are specifically processed by Cas and Cse proteins associated with that locus.[2] The type II system from *S. pyogenes* that encodes Cas9 is the most well studied.

CRISPR/Cas9 has been used extensively in research as an efficient system to alter DNA at specific locations in the genome. Synthetic crRNA and tracRNA can be designed to lead the Cas9 endonuclease to the site of

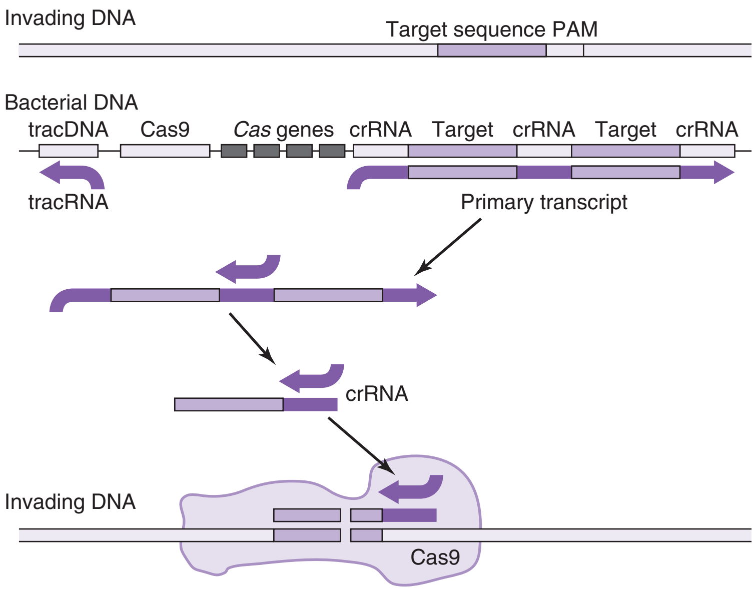


FIGURE 5.4 Clustered regularly interspaced short palindromic repeats (CRISPR) is a protective system that uses the invading DNA sequences to target itself. The CRISPR locus consists of regularly spaced short palindromic repeats. Foreign DNA sequences are incorporated into the bacterial genome at the CRISPR repeat loci, interspersed with target (invader) DNA, tracRNA the gene, and the CRISPR-associated (Cas) operon. CRISPR loci are then transcribed and processed into crRNA. crRNA and tracRNA combine with the Cas enzyme to find homology with invading DNA adjacent to the PAM sequence, at which site the enzyme will cut the invading DNA.

Historical Highlights

Before CRISPR was described, targeted genome editing was performed in vitro with transcription activator–like effector nucleases (TALENs) and zinc finger nucleases (ZFNs).[3] These engineered nucleases contained a customized target-sequence-specific DNA-binding domain fused to a nuclease that would cut DNA at any sequence. The nucleases made double-strand breaks (DSBs) into targeted DNA sites. Designing and engineering the proteins, however, was technically demanding and time-consuming, limiting their use in many applications.

Advanced Concepts

Targeted genome editing is the modification of a sequence of interest in living cells or organisms. Edits of early-stage embryos or even zygotes modify the genome in all the cells of an organism. The first disease gene repair was performed in mice where a gene mutation that caused lens clouding (cataracts) in mice was targeted with a dominant normal gene using the CRISPR/Cas9 system.[6] Modification of autologous hematopoietic stem cells using CRISPR technology is a promising approach to the treatment of blood disorders currently treated with bone marrow transplants.[7]

choice, providing the specificity of restriction enzymes with the versatility of guiding cuts to desired sequence sites. Once the DNA is cleaved, the cell will repair the break by homologous recombination with a synthetic donor template providing any desired sequence changes. CRISPR RNA can also lead activators or repressors instead of Cas9 to gene-promoter sites and affect gene transcription. Specific regions can be visualized by bringing reporter molecules, such as green fluorescent protein, in place of Cas9.[4,5]

HYBRIDIZATION TECHNOLOGIES

Procedures performed in the clinical molecular laboratory are aimed at specific targets in genomic DNA. This requires visualization or detection of a particular gene or region of DNA in the backdrop of all other genes. There are several ways to find a target region of DNA, some of which require cloning of the region of interest. Early cloning methods were complex, requiring screening of thousands of plasmids into which DNA regions were

TABLE 5.1 Hybridization Technologies

Hybridization Method	Target	Probe	Purpose
Southern blot	DNA	Nucleic acid	Gene structure
Northern blot	RNA	Nucleic acid	Transcript structure, processing, gene expression
Western blot	Protein	Protein	Protein processing, gene expression
Southwestern blot	Protein	DNA	DNA-binding proteins, gene regulation
Eastern blot	Protein	Protein	Modification of western blot using enzymatic detection (PathHunter); also, detection of specific agriculturally important proteins[28]
Far-eastern blot[29,30]	Lipids	(None)	Transfer of high-performance liquid chromatography (HPLC)-separated lipids to polyvinyl difluoride (PVDF) membranes for analysis by mass spectrometry

randomly inserted. The first method for molecular analysis of specific DNA sites within a complex background without cloning that region was the Southern blot. Modifications of the Southern blot are applied to the analysis of RNA, proteins, and lipids in order to study gene expression, regulation, and protein modifications (see Table 5.1).

Southern Blots

The **Southern blot** is named for Edwin Southern, who first reported the procedure.[8] In the Southern blot, genomic DNA is isolated and cut with restriction enzymes. The fragments are separated by gel electrophoresis, denatured, and then transferred to a solid support such as **nitrocellulose.** In the final steps of the procedure, the DNA fragments are exposed to a labeled **probe** (complementary DNA or RNA) that is complementary to the region of interest. The signal of the probe is detected to indicate the presence or absence (lack of signal) of the sequence in question. The original method entailed hybridization of a radioactively labeled probe to detect the DNA region to be analyzed. As long as there is a probe of known identity, this procedure can analyze any gene or gene region in prokaryotes or eukaryotes at the molecular level. Newer methods have replaced PCR for many applications; however, Southern blots are still applied to the characterization of large regions (10 kb to more than 100 kb). The following sections describe the parts of the Southern blot procedure in detail and discuss modifications of the procedure in order to analyze RNA and protein.

Restriction Enzyme Cutting and Resolution

The first step in the Southern blot procedure is digestion of test DNA with restriction enzymes. The choice of enzymes used will depend on the genetic locus and the application. For routine laboratory tests, restriction maps of the target DNA regions will have previously been determined, and the appropriate enzymes will have been recommended. For other methods, such as typing of unknown organisms or cloning, several enzymes may be tested to find those that will be most informative.

Ten to 50 μg of high-quality (intact) genomic DNA is used for each restriction enzyme digestion for Southern analysis. More or less DNA may be required depending on the sensitivity of the detection system, the volume and configurations of wells, and the abundance of the target DNA. In the clinical laboratory, specimen availability may limit the amount of DNA available for testing. The DNA is mixed with the appropriate restriction enzyme and buffer. Restriction enzymes are supplied with 10× buffers that are diluted $^1/_{10}$ into the final reaction volume. If more than one enzyme is to be used, each sample must be digested separately for each enzyme and buffer.

Digestion is carried out for an extended time (3 hours or more) to allow complete cutting of all sites in the DNA sample. High-specific-activity enzymes are also used to ensure complete cutting of every site. Incomplete cutting will result in anomalous patterns, complicating interpretation of the Southern blot data. The fragments resulting from the restriction digestions are resolved by gel electrophoresis. The percentage and nature of the gel will depend on the size of the DNA region to be analyzed (see Chapter 4, Tables 4.1 and 4.2). As with all electrophoresis, a molecular-weight standard should be run with the test samples. Large fragments require longer runs at low voltage to get the best resolution. For example, 10,000- to 20,000-bp fragments are resolved in 0.7% agarose at 20 amperes for 16 hours.

After electrophoresis, it is important to observe the cut DNA. Figure 5.5 shows a 0.7% agarose gel stained with ethidium bromide and illuminated by ultraviolet (UV) light. Genomic DNA cut with restriction enzymes should produce a smear representing billions of fragments of all sizes released by enzyme digestion. The brightness of the DNA smears should be similar from lane to lane, ensuring that equal amounts of DNA were added to all lanes. A large aggregate of DNA near the top of the lane indicates that the restriction enzyme activity was incomplete, preventing size analysis of the sample. A smear located primarily in the lower region of the lane is a sign that the DNA is degraded. Uncut or degraded DNA will prevent accurate analysis. Repeat of the restriction digest will be required if uncut DNA is present. If an impurity in the DNA results in resistance to restriction digestion or degradation of the DNA, re-isolation or further purification of the DNA will be required.

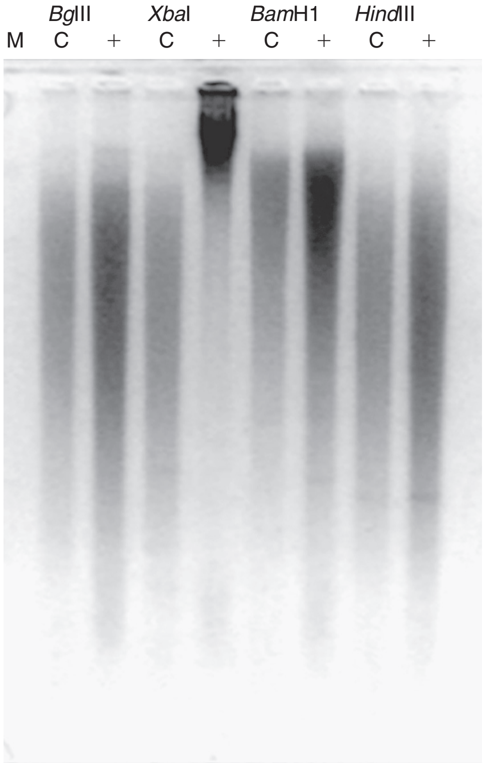


FIGURE 5.5 Properly restricted genomic DNA will produce a smear of fragments along the agarose gel lane ranging in size depending on the frequency of the particular enzyme restriction sites in the DNA. Among these are the fragments coming from the region under study. After probing, only those fragments will be visible.

Preparation of Resolved DNA for Blotting (Transfer)

The goal of the Southern blot procedure is to analyze a specific region of the sample DNA. The restriction fragments containing the target sequence to be analyzed are obviously not distinguishable in the collection of other fragments that do not have the target sequence. Target fragments can be detected by hybridization with a complementary sequence of single-stranded DNA or RNA labeled with a detectable signal. To achieve optimal hydrogen bonding between the probe and its complementary sequence in the resolved sample DNA, the double-stranded DNA fragments in the gel must be separated into single strands (denatured) and transferred to a nitrocellulose membrane.

Depurination

Before moving the DNA fragments from the gel to the membrane for blotting, the double-stranded DNA fragments are denatured as the DNA remains in place in the gel. Although short fragments can be denatured directly as described in the following section, larger fragments (greater than 500 bp) are more efficiently denatured if they are depurinated before denaturation (Fig. 5.6). Therefore, for large fragments, the gel is first soaked in dilute hydrogen chloride (HCl) solution, a process that removes purine bases from the sugar-phosphate backbone. This will “loosen up” the larger fragments for more complete denaturation.

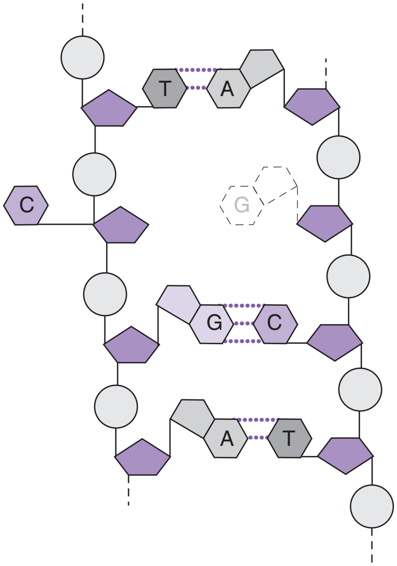


FIGURE 5.6 An apurinic site in double-stranded DNA. Loss of the guanine leaves an open site but does not break the sugar-phosphate backbone of the DNA.

Denaturation

DNA is denatured by exposing the gel to a strong base such as sodium hydroxide (NaOH). NaOH promotes breakage of the hydrogen bonds holding the DNA strands to one another. The resulting single strands are then available to hydrogen bond with the single-stranded probe. Further, the single-stranded DNA will bind more tightly than double-stranded DNA to the nitrocellulose membrane upon transfer.

Blotting (Transfer)

Before exposing the denatured sample DNA to the probe, the DNA is transferred, or blotted, to a solid substrate that will facilitate probe binding and signal detection. This substrate is usually nitrocellulose, nylon, or cellulose modified with a diethyl aminoethyl, or a carboxymethyl (CM) chemical group. Membranes of another type, polyvinyl difluoride (PVDF), are used for immobilizing proteins for probing with antibodies (western blots).

Membrane Types

Single-stranded DNA avidly binds to nitrocellulose membranes with a noncovalent, but irreversible, connection. The binding interaction is hydrophobic and electrostatic between the negatively charged DNA and the positive charges on the membrane. Nitrocellulose-based membranes bind 70 to 150 μg of nucleic acid per square centimeter. Membrane pore sizes (0.05 to 0.45 μm) are suitable for DNA fragments from a few hundred bases up to those greater than 20,000 bp in length.

Advanced Concepts

Treatment of DNA with dilute (0.1 to 0.25 mM) hydrochloric acid results in hydrolysis of the glycosidic bonds between purine bases and the sugar of the nucleotides. This loss of purines (adenines and guanines) from the sugar-phosphate backbone of DNA leaves apurinic sites (see Fig. 5.6). The DNA backbone remains intact and holds the rest of the bases in linear order. Removal of some of the purine bases facilitates the subsequent breaking of hydrogen bonds between the two strands of the DNA during the denaturation step in Southern blotting.

Pure nitrocellulose has a high binding capacity for proteins as well as nucleic acids. It is the most versatile medium for molecular transfer applications. It is also compatible with different transfer buffers and detection systems. Nitrocellulose is not as sturdy as other media and becomes brittle upon drying. Because all genomic fragments are permanently bound to the membrane, the hybridized probe can be stripped, and a second probe can be hybridized to the same membrane. Reinforced nitrocellulose is more appropriate for applications where multiple probings may be necessary. Mechanically stable membranes can be formulated with a net neutral charge to decrease nonspecific binding. These membranes have a very high binding capacity (less than 400 μg/cm^2), which increases the test sensitivity. A covalent attachment of nucleic acid to these membranes is achieved by exposure of the DNA on the membrane to UV-light cross-linking.

Membranes with a positive charge more effectively bind small fragments of DNA. These membranes, however, are more likely to retain protein or other contaminants that will contribute to background noise after the membrane is probed.

Before transfer of the sample, membranes are moistened by floating them on the surface of the transfer buffer. Any dry spots (areas where the membrane does not properly hydrate) will remain white while the rest of the membrane darkens with buffer. If the membrane does not hydrate evenly, dry spots will inhibit binding of the sample.

Advanced Concepts

Binding of single-stranded DNA to nitrocellulose does not prevent hydrogen-bond formation of the immobilized DNA with complementary sequences. The bond between the membrane and the DNA is much stronger than the hydrogen bonds that hold complementary strands together. This allows for removal of probes and re-probing of different target regions.

Transfer Methods

Transfer of nucleic acid to protein is performed in several ways. The goal is to move the DNA from the gel to a membrane substrate for probing. The membrane must be equilibrated in the transfer buffer before coming into contact with the DNA in the gel. Membranes should be handled carefully, preferably with powder-free gloves, avoiding folding or creasing of the membrane. The original method developed by Southern used capillary transfer (Fig. 5.7). For capillary transfer, the gel is placed on top of a reservoir of buffer, which can be a shallow container or filter papers soaked in high-salt buffer, for example, 10× saline sodium citrate (10X SSC: 1.5 M NaCl, 0.15 M Na citrate) or commercially available transfer buffers. The nitrocellulose membrane is placed directly on the gel, and dry absorbent filter paper or paper towels are stacked on top of the membrane. The buffer will move by capillary action from the lower reservoir to the dry material on top of the gel. The movement of the buffer transversely through the gel will carry the denatured DNA out of the gel. When the DNA contacts the nitrocellulose membrane, the DNA will bind to it.

Advanced Concepts

Diethylaminoethyl (DEAE)-conjugated cellulose effectively binds nucleic acids and negatively charged proteins. PVDF and charged carboxymethyl cellulose membranes are used only for protein (western) blotting. These membranes bind nucleic acid and proteins by hydrophobic and ionic interactions with a binding capacity of 20 to 40 μg/cm^2 to 150 μg/cm^2 for PVDF. Modifications of PVDF are designed to optimize use in a variety of test and sample types. These include small-pore-size membranes for use with small proteins and membranes optimized for fluorescent detection.

Capillary transfer is simple and relatively inexpensive. No instruments are required. The transfer, however, can be less than optimal, especially with large gels. Bubbles, salt crystals, or other particles between the membrane and the gel can cause loss of information or staining artifacts. The procedure is also slow, taking from a few hours to overnight for large fragments.

A second method, called **electrophoretic transfer,** uses electric current to move the DNA from the gel to the membrane (Fig. 5.8). This system utilizes electrodes attached to membranes above (anode) and below (cathode) the gel. The current carries the DNA transversely from the gel to the membrane. Electrophoretic

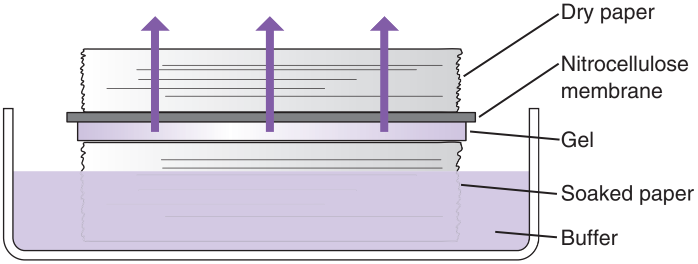


FIGURE 5.7 Capillary transfer. Driven by capillary movement of buffer from the soaked paper to the dry paper, denatured DNA moves from the gel to the membrane. The DNA will adhere to the membrane, which will be subsequently exposed to the probe.

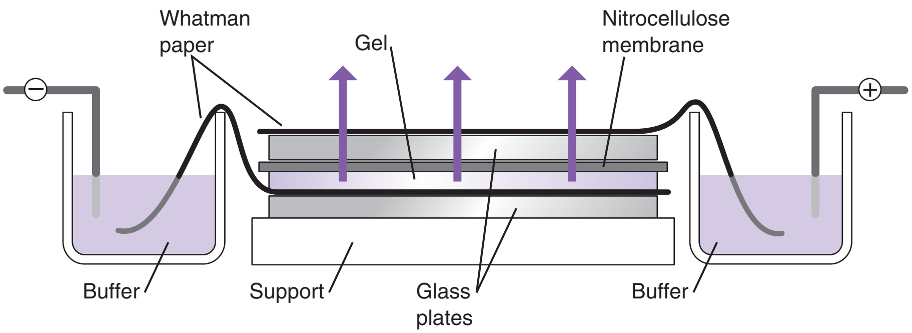
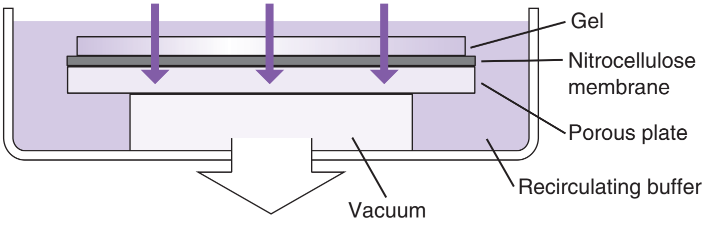


FIGURE 5.8 Electrophoretic transfer. This system uses electric current to move the DNA transversely through the gel to the membrane. This type of transfer is used mostly for small fragments or proteins.

FIGURE 5.9 Vacuum transfer. This system uses suction and buffer recirculation to move the DNA out of the gel and onto the membrane. Vacuum transfer is generally faster than capillary transfer for large DNA fragments; however, unlike capillary transfer, vacuum transfer requires specialized equipment.



transfer is performed in a “tank” or by a “semidry” approach. In the tank method, the electrodes transfer current through the gel and membrane through electrophoresis buffer, as shown in Figure 5.8. In the semidry method, the electrodes contact the gel-membrane sandwich directly, requiring only enough buffer to soak the gel and membrane. The tank electrophoretic transfer is preferred for large proteins resolved on acrylamide gels, whereas the semidry method is frequently used for small proteins.

Vacuum transfer is a third method of DNA blotting (Fig. 5.9). This blotting technique uses suction to move the DNA from the gel to the membrane in a recirculating buffer. Like electrophoretic transfer, this method transfers the DNA more rapidly than capillary transfer—in 2 to 3 hours rather than overnight. Also, it avoids discontinuous transfer due to air trapped between the membrane and the gel. One disadvantage of the second and third methods is the expense and maintenance of the electrophoresis and vacuum equipment.

After transfer, the cut, denatured DNA is avidly bound to the membrane. The DNA can be permanently immobilized to the membrane by baking in a vacuum oven (80°C, 30 to 60 minutes) or by UV cross-linking, that is, covalently attaching the DNA to the nitrocellulose using UV-light energy. Baking or cross-linking covalently attaches the DNA to the membrane and prevents the DNA fragments from washing away or moving on the membrane during extended procedures or when sequential probings are to be done.

PROBE HYBRIDIZATION

Following immobilization of the DNA, a prehybridization step is required to prevent the probe from binding to nonspecific sites on the membrane surface, which may cause background noise. **Prehybridization** involves incubating the membrane in the same buffer in which the probe will subsequently be introduced or in a specially formulated prehybridization buffer solution. At this point, the buffer does not contain probe. Prehybridization buffer consists of such blocking agents as Denhardt solution (Ficoll, polyvinyl pyrrolidine, bovine serum albumin) and salmon sperm DNA. Sodium dodecyl sulfate (SDS, 0.01%) may also be included, along with formamide, the latter especially for RNA probes. The membrane is exposed to the prehybridization buffer at the optimal hybridization temperature for 30 minutes to several hours, depending on the specifications in the protocol. At this stage, the sample is ready for hybridization with the probe, which will allow visualization of the specific gene or region of interest.

Northern Blots

The **northern blot** technique, a modification of the Southern blot technique, was designed to investigate RNA structure and quantity. Although most northern analyses were performed to investigate levels of gene expression (transcription from DNA) and stability, the analyses were also used to investigate RNA structural abnormalities resulting from aberrations in synthesis or processing, such as alternative splicing. Splicing abnormalities are responsible for a number of diseases, such as beta-thalassemias and familial isolated growth hormone deficiency. Analysis of RNA structure and quantity indirectly reveals mutations in the regulatory or splicing signals in DNA.

To perform a northern blot, nucleic acid isolation methods for RNA are used. An RNase-free environment must always be maintained for RNA preparation. After isolation and quantification of RNA, the samples (up to approximately 30 μg total RNA or 0.5 to 3.0 μg polyA RNA, depending on the relative abundance of the transcript under study) are applied directly to agarose gels. Agarose concentrations of 0.8% to 1.5% are usually employed. Polyacrylamide gels may also be used, especially for smaller transcripts—for instance, for analysis of viral gene expression. Gel electrophoresis of RNA is carried out under denaturing conditions for accurate transcript size assessment. Complete denaturation is also required for efficient transfer of the RNA from the gel to the membrane, as with the transfer of DNA in the Southern blot. Because the denaturation is maintained during electrophoresis, a separate denaturation step is not required for northern blots. After electrophoresis, representative lanes can be cut from the gel, soaked in ammonium acetate to remove the denaturant, and stained with acridine orange or ethidium bromide to assess quality and equivalent sample loading.

Denaturant such as formaldehyde must be removed from the gel before transfer because it inhibits binding of the RNA to nitrocellulose. This is accomplished by rinsing the gel in deionized water. RNA is transferred in 10× or 20× SSC or 10× SSPE (1.8 M NaCl, 0.1 M sodium phosphate, pH 7.7, 10 mM EDTA) to nitrocellulose as described previously for DNA. For small transcripts (500 bases or less), 20× SSC should be used. The blotting procedure for RNA in the northern blot is performed in 20× SSC, similar to the procedure for DNA transfer in the Southern blot. Prehybridization and hybridization in formamide/SSC/SDS prehybridization/hybridization buffers are also performed as with Southern blot. If the RNA has been denatured in glyoxal, the membrane must be soaked in warm Tris buffer (65°C) to remove any residual denaturant immediately before prehybridization.

Although it is still used in some research applications, the northern blot has been mostly replaced by more efficient technologies with lower sample demands. Genomic technologies, such as arrays (described later in this chapter), provide more comprehensive information with less demand on RNA length.

Western Blots

Another modification of the Southern blot is the **western blot.**[9] The immobilized target for a western blot is protein. There are many variations on western blots. Generally, serum, cell lysate, or extract is separated on SDS-polyacrylamide gels (SDS-PAGE) or isoelectric focusing gels (IEF). The former resolves proteins according to molecular weight, and the latter resolves proteins according to charge. Dithiothreitol or 2-mercaptoethanol may also be used to separate proteins into subunits. Polyacrylamide concentrations vary from 5% to 20%. Depending on the complexity of the protein and the quantity of the target protein, 1 to 50 μg of protein is loaded per well. Before loading, the sample is treated with denaturant, such as mixing 1:1 with 0.04 M Tris HCl, pH 6.8, 0.1% SDS. The accuracy and sensitivity of the separation can be enhanced by using a combination of IEF gels followed by SDS-PAGE or by using two-dimensional gel electrophoresis. Prestained molecular-weight standards are run with the samples to orient the membrane after transfer and to approximate the sizes of the proteins after probing. Standards ranging from 11,700 d (cytochrome C) to 205,000 d (myosin) are commercially available.

The gel system used may affect subsequent probing of proteins with antibodies. Specifically, **denaturing gels** could affect **epitopes** (antigenic sites on the protein) such that they will not bind with the labeled antibodies. Gel pretreatment with mild buffers, such as 20% glycerol in 50 mM Tris-HCl, pH 7.4, can renature proteins before transfer.

After electrophoresis, proteins are blotted to membranes by capillary or electrophoretic transfer. Nitrocellulose has a high affinity for proteins and is easily treated with detergent (0.1% Tween 20 in 0.05 M Tris and 0.15 M sodium chloride, pH 7.6) with 5% dry milk to prevent binding of the **primary antibody** probe to the membrane itself **(blocking)** before hybridization. Binding of proteins to nitrocellulose is probably hydrophobic because nonionic detergents can remove proteins from the membrane. Other membrane types used for protein blotting are PVDF and anion (DEAE) or cation (CM) exchange cellulose. After incubation with the primary antibody for 12 to 16 hours, the blot is washed in the same buffer (without dry milk) and incubated with the secondary antibody conjugated with enzyme. The blot is washed again to remove excess secondary antibody conjugate, and the chemiluminescent or color signal is developed with the addition of substrate.

Advanced Concepts

A frequent application of the western blot method is confirmation of enzyme-linked immunoassay (ELISA) results for human immunodeficiency virus (HIV) and hepatitis C virus (HCV), among other microbes. In this procedure, known viral proteins are separated by electrophoresis and transferred and bound to a nitrocellulose membrane. The patient's serum is overlaid on the membrane, and antibodies with specificity to viral proteins bind to their corresponding protein antigens. Unbound patient antibodies are washed off, and binding of antibodies is detected by adding a labeled antihuman immunoglobulin antibody. If viral antibodies are present in the patient's serum, they can be detected with antihuman antibody probes appearing as a dark band on the blot corresponding to the specific HIV protein to which the antibody is specific.

PROBES

The probe for Southern and northern blots is a single-stranded fragment of nucleic acid attached to a signal-producing moiety. The purpose of the probe is to identify one or more sequences of interest within a large amount of nucleic acid. The probe therefore should hybridize specifically with the target DNA or RNA that is to be analyzed. The probe can be RNA, denatured DNA, or other modified nucleic acids. **Peptide nucleic acids (PNAs)** and **locked nucleic acids** have also been used as probes. These structures contain normal nitrogen bases that can hybridize with complementary DNA or RNA, but the bases are connected by backbones different from the natural phosphodiester backbone of DNA and RNA. These modified nucleic acids are resistant to nuclease degradation and, because of a reduced negative charge on their backbone, can hybridize more readily to target DNA or RNA.

Probes for western blots are specific binding proteins or antibodies. A labeled **secondary antibody** directed against the primary binding protein is then used for the visualization of the protein band of interest.

DNA Probes

DNA probes are created in several ways. In early methods, a fragment of the gene to be analyzed was cloned on a bacterial plasmid and then isolated by restriction enzyme digestion and gel purification. The fragment, after labeling (see following discussion) and denaturation, was then used in Southern or northern blot procedures.

Other sources of DNA probes include the isolation of a sequence of interest from viral genomes and in vitro organic synthesis of nucleic acid of a predetermined sequence. The latter is used only for short, oligomeric probes. Probes may also be synthesized using the polymerase chain reaction (PCR).

The length of the probe will, in part, determine the specificity of the hybridization reaction. Probe lengths range from tens to thousands of base pairs. In an analysis of the entire genome in a Southern blot, longer probes are more specific for a DNA region because they must distinguish among many closely similar sequences. Shorter probes are not usually used in Southern blots because short sequences are more likely to be found in multiple locations in the genome, resulting in high background binding to sequences not related to the target region of interest.

The probe is constructed so that it has a complementary sequence to the targeted gene. In order to bind to the probe, then, the target nucleic acid has to contain the sequence of interest. Properly prepared and stored DNA probes are relatively stable. Double-stranded DNA probes must be denatured before use. This is usually accomplished by heating the probe (e.g., 95°C, 10 to 15 min) in hybridization solution or treating with 50% formamide/2× SSC at a lower temperature for a shorter time (e.g., 75°C, 5 to 6 min).

RNA Probes

RNA probes are often made by transcription from a synthetic DNA template in vitro. These probes are similar to DNA probes with equal or greater binding affinity to complementary sequences. Because RNA and DNA form a stronger helix than DNA/DNA, the RNA probes may offer more sensitivity than DNA probes in the Southern blot.

RNA probes can be synthesized directly from a plasmid template or from template DNA produced by PCR. Predesigned systems commercially available for this purpose include plasmid vector DNA containing a binding site for RNA polymerase (promoter), a cloning site for the sequences of interest, and a DNA-dependent RNA polymerase from *Salmonella* bacteriophage SP6 or *Escherichia coli* bacteriophage T3 or T7. DNA sequences complementary to the RNA transcript to be analyzed are cloned into the plasmid vector using restriction enzymes. The recombinant vector containing the gene of interest is then linearized, and the RNA probe is transcribed in vitro from the promoter. The coding strand transcript can be used for Southern blots. For northern blots, the antisense transcript (complementary to the coding strand transcript on the blot) is generated by inverted placement of the fragment so that the promoter starts at the end of the gene. Either coding or complementary RNA will hybridize to a double-stranded DNA target. RNA probes for northern blots must be designed so that the probe is complementary to the target sequence. A probe of identical sequence to the target RNA (coding sequence) will not hybridize.

RNA probes are labeled to produce a signal by incorporating a radioactive or modified nucleotide during the in vitro transcription process. Labeling during synthesis provides RNA probes with a high specific activity (signal to micrograms of probe) that increases the sensitivity of the probe. To avoid background noise, some protocols include digestion of nonhybridized probe, using a specific RNase such as RNase A, after hybridization is complete.

RNA probes are generally less stable than DNA probes and cannot be stored for long periods. Synthesis of an RNA probe by transcription from a stored template is relatively simple and is easily performed within a few days of use. The DNA template can be removed from the probe by treatment with RNase-free DNase. Although RNA is already single stranded, denaturation before use is recommended in order to eliminate the secondary structure internal to the RNA molecule.

Other Nucleic Acid Probe Types

Peptide nucleic acid, locked nucleic acid, and unlocked nucleic acid probes (Figs. 5.10 and 5.11) are synthesized using chemical methods.[10] These modified nucleic acids have the advantage of being resistant to nucleases that degrade DNA and RNA by breaking the phosphodiester backbone. Further, the negative charge of the phosphodiester backbones of DNA and RNA counteract hydrogen bonding between the bases of the probe and target sequences. Structures such as peptide nucleic acid (PNA) that do not have a negative charge hybridize more efficiently. Unlocked nucleic acids (UNA) lack the C2′–C3′ ribose sugar bond found in ribonucleoside, resulting in high flexibility. Incorporation of UNA nucleotides destabilizes the double helix. UNA can therefore “fine-tune” the hybridization of probes to targets, especially in short sequences.[11]

Protein Probes

Western blot protein probes are antibodies that bind specifically to the immobilized target protein. Polyclonal or monoclonal antibodies are used for this purpose. **Polyclonal antibodies** are products of a generalized response

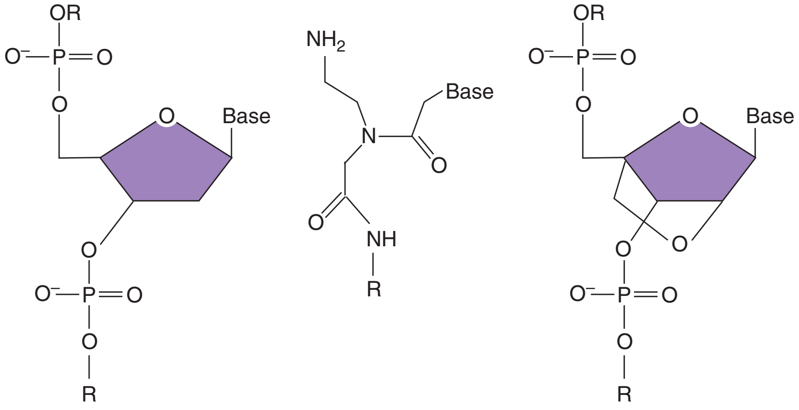

OR
O⁻
P
O
Base
NH2
N
NH
R

FIGURE 5.10 Peptide nucleic acids have the phosphodiester bond *(left)* replaced with carbon-nitrogen peptide bonds *(center)*. Locked nucleic acids are bicyclic nucleoside monomers where the ribose sugar contains a methylene link between its 2′ oxygen and 4′ carbon atoms *(right)*.

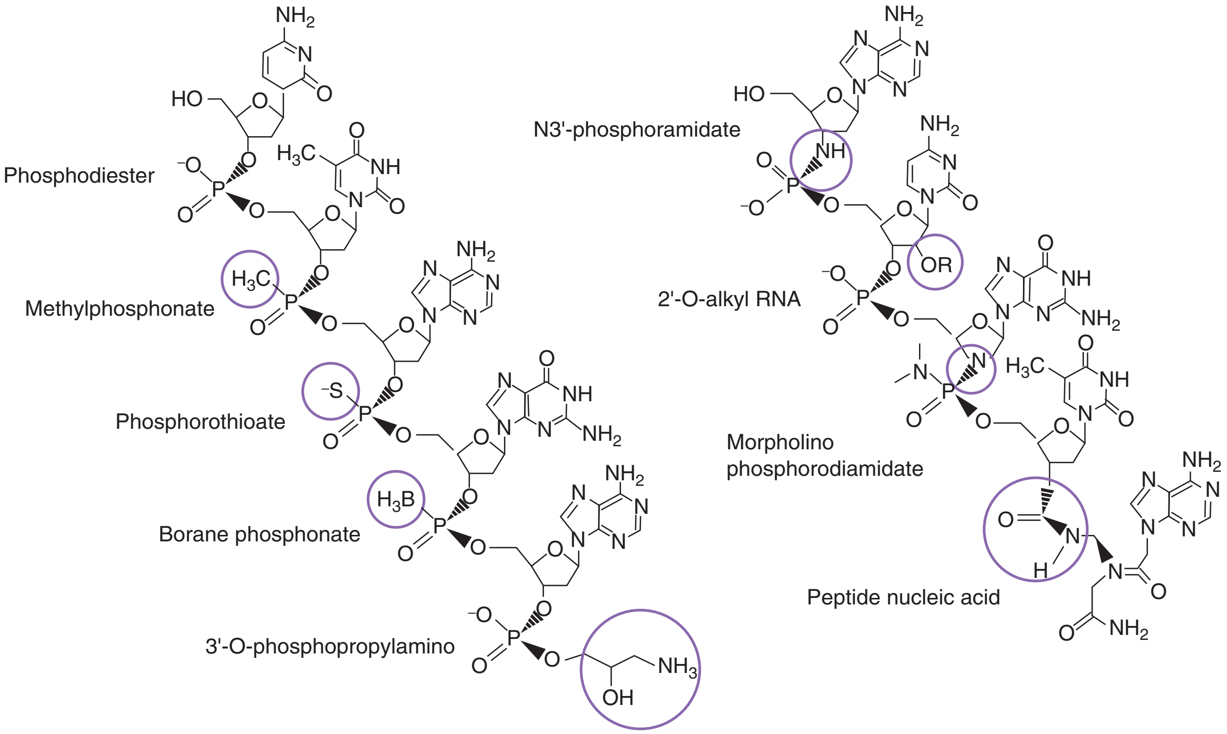

Phosphodiester
Methylphosphonate
Phosphorothioate
Borane phosphonate
3'-O-phosphopropylamino
N3'-phosphoramidate
2'-O-alkyl RNA
Morpholino phosphorodiamidate
Peptide nucleic acid

FIGURE 5.11 Modifications of the phosphodiester backbone of nucleic acids include substitution on the alpha-phosphate group with alkyl, sulfur, or other groups. The ribose may also have modifications with nitrogen or alkyl groups or replacement with peptide rather than phosphodiester bonds *(right)*.

to a specific antigen, usually a peptide or protein. Small molecules **(haptens)** attached to protein carriers, carbohydrates, nucleic acids, and even to whole cells and tissue extracts can be used to generate an antibody response. Adjuvants, such as Freund's adjuvant, enhance the antibody titer by slowing the degradation of the protein and lengthening the time the immune system is exposed to the stimulating antigen. Specific immunoglobulins are subsequently isolated from sera by affinity chromatography.

Polyclonal antibodies are comprised of a mixture of immunoglobulins directed at more than one epitope (molecular structure) on the antigen. **Monoclonal antibodies** are more difficult to produce. Kohler and Milstein first demonstrated that spleen cells from immunized mice could be fused with mouse myeloma cells to form hybrid cells **(hybridomas)** that could grow in culture and secrete antibodies.[12] By cloning the hybridomas (growing small cultures from single cells), preparations of specific antibodies could be produced continuously. The clones could then be screened for antibodies that best react with the target antigen. The monoclonal antibodies are isolated from cell culture fluid; higher titers of antibodies are obtained by inoculating the antibody-producing hybridoma into mice and collecting the peritoneal fluid. The monoclonal antibody is then isolated by chromatography.

Polyclonal antibodies are useful for immunoprecipitation methods and for western blots. With their greater specificity, monoclonal antibodies can be used for almost any procedure. In western blot technology, polyclonal antibodies can give a more robust signal, especially if the target epitopes are partially lost during electrophoresis and transfer. Monoclonal antibodies are more specific and may give less background noise; however, if the targeted epitope is lost, these antibodies do not bind, and no signal is generated. The dilution of primary antibody can range from 1/100 to 1/100,000, depending on the sensitivity of the detection system.

Advanced Concepts

Alternative nucleic acids are not only useful in the laboratory, but they are also potentially valuable in the clinic. Several structures have been proposed for use in antisense gene therapy (Fig. 5.11). Introduction of sequences complementary to messenger RNA of a gene (antisense sequences) will prevent translation of that mRNA and expression of that gene. If this could be achieved in whole organisms, selected aberrantly expressed genes or even viral genes could be turned off. One drawback of this technology is the degradation of natural RNA and DNA by intracellular nucleases. The nuclease-resistant structures are more stable and available to hybridize to the target mRNA.

Probe Labeling

In order to visualize the probe bound to target fragments on a membrane, the probe must be labeled and generate a detectable signal. In the past, classical Southern analysis methods called for radioactive labeling with ^{32}P. This labeling was achieved by introduction of nucleotides containing radioactive phosphorus to the probe. Today, many medical laboratories now use nonradioactive labeling to avoid the hazard and expense of working with radiation. Nonradioactive labeling methods are based on indirect detection of a tagged nucleotide incorporated in or added to the probe. The two most commonly used nonradioactive tags are **biotin** and **digoxigenin** (Fig. 5.12), either of which can be attached covalently to a nucleotide triphosphate, usually UTP or CTP.

There are three basic methods that are used to label a DNA probe: end labeling, nick translation, and random priming. In **end labeling,** labeled nucleotides are added to the end of the probe using terminal transferase or T4 polynucleotide kinase. In **nick translation,** the labeled nucleotides are incorporated into single-stranded breaks, or nicks, that are substrates for nucleotide addition by DNA polymerase. The polymerase uses the intact complementary strand for a template and displaces the previously hybridized strand as it extends the nick. **Random priming** generates new single-stranded versions of the probe with the incorporation of the labeled nucleotides. The synthesis of these new strands is primed by oligomers of random sequences that are 6 to 10 bases in length. These short sequences will, at some frequency, complement sequences in the denatured probe and prime synthesis of copies of the probe sequences with incorporated labeled nucleotides.

RNA probes are transcribed from cloned DNA or amplified DNA. These probes are labeled during their synthesis with radioactive, biotinylated, or digoxigenin-tagged nucleotides. Unlike double-stranded complementary DNA probes and targets that contain both strands of the complementary sequences, RNA probes are single stranded, with only one strand of the complementary sequence represented.

Nucleic Acid Probe Design

The most critical parts of any hybridization procedure are the design and optimal hybridization of the probe,

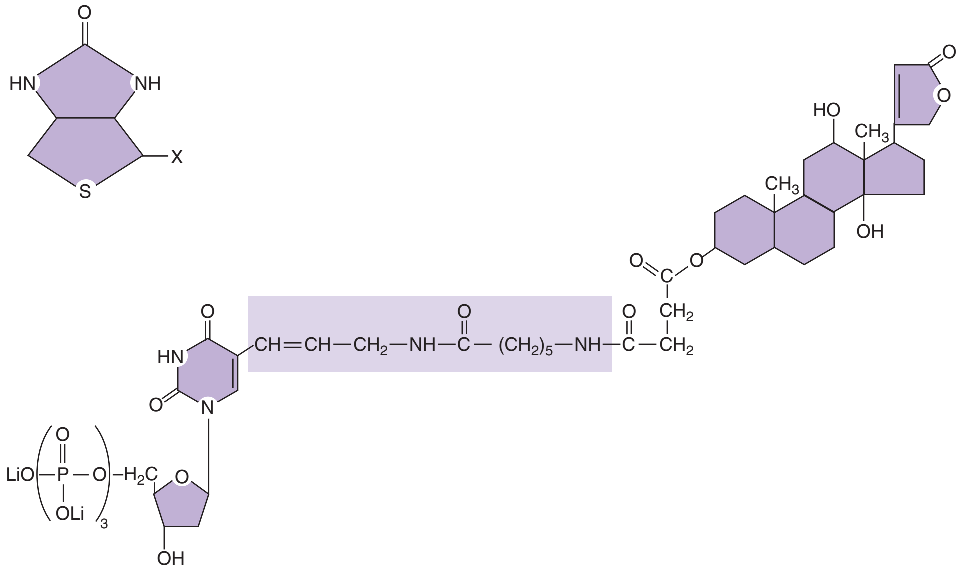


FIGURE 5.12 Biotin *(top)* has a variable side chain (X). The polycyclic digoxigenin *(bottom)* is shown covalently attached to UTP (dig-11-UTP). This molecule can be covalently attached or incorporated into DNA or RNA to make a labeled probe.

both of which determine the specificity of the results. With nucleic acids, the more optimal the hybridization conditions for a probe/target interaction, the more specific the probe. Longer probes (500 to 5,000 bp) offer greater specificity with decreased background noise because they are less affected by point mutations or polymorphisms within the target sequence or within the probe itself. Long probes, however, may be difficult or expensive to synthesize.

Shorter probes (less than 500 bp) are less specific than longer ones in Southern blotting applications. A short sequence has a higher chance of being repeated randomly in unrelated regions of the genome. Short probes are ideal, however, for mutation analysis because their binding affinity is sensitive to single-base-pair changes within a target binding sequence.

Advanced Concepts

The protein probes used in western blot applications may be labeled with ^{35}S for radioactive detection. For nonradioactive detection, western protein probes are covalently bound to an enzyme, usually horseradish peroxidase (HRP) or alkaline phosphatase. When exposed to a light- or color-generating substrate, the enzyme will produce a detectable signal on the membrane or on an autoradiogram. Unconjugated antibodies are detected after binding with a conjugated secondary antibody to the primary probe, such as mouse antihuman or rabbit antimouse antibodies (Fig. 5.13). The secondary antibodies will recognize any primary antibody by targeting the FC region.

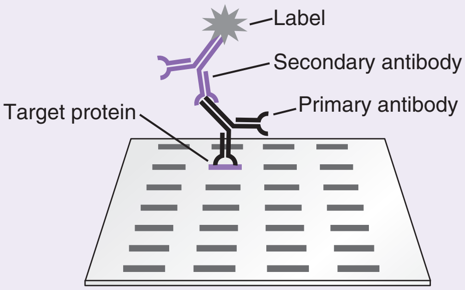


FIGURE 5.13 Probe binding to western blots may include an unlabeled primary antibody that is bound by a secondary antibody carrying a label for detection.

The probe sequence can affect its binding performance. A probe with internal complementary sequences will fold and hybridize with itself, which will compete with hybridization to the intended target. The probe folding or secondary structure is especially strong in sequences with high GC content, decreasing the binding efficiency to the target sequence and, therefore, the test sensitivity.

HYBRIDIZATION CONDITIONS, STRINGENCY

Southern blot and northern blot probing conditions must be empirically optimized for each nucleic acid target. Stringency is the combination of conditions under which the target is exposed to the probe. Conditions of high stringency are more demanding of probe/target complementarity and length. Low-stringency conditions are more forgiving. If conditions of stringency are set too high, the probe will not bind to its target. If conditions are set too low, the probe will bind unrelated targets, complicating the interpretation of the final results.

Several factors affect stringency. These include the temperature of hybridization; the salt concentration of the hybridization buffer; and the concentration of denaturant, such as formamide, in the buffer. The length and nature of the probe sequence can also influence the level of stringency. A long probe or one with a higher percentage of G and C bases will bind under more stringent conditions than a short probe or one with greater numbers of A and T bases will bind. The ideal hybridization conditions are estimated from the calculation of the **melting temperature,** or T_m, of the probe sequence. The T_m is a way to express the amount of energy required to separate the hybridized strands of a given sequence (Fig. 5.14). At the T_m, half of the sequence is double stranded, and half is single stranded. One formula for the T_m of a double-stranded DNA sequence in solution is

$$T_m = 81.5°C + 16.6 \log M + 0.41(\%G + C) - 0.61(\% \text{ formamide}) - (600/n)$$

where M = sodium concentration in mol/L, and n = number of base pairs in the shortest duplex.

RNA:RNA hybrids are more stable than DNA:DNA hybrids due to less constraint by the RNA phosphodiester backbone. DNA:RNA hybrids have intermediate affinity. The formulae for RNA:RNA or DNA:RNA hybrids, therefore, are slightly different from the formula for a DNA:DNA helix. RNA:DNA hybrids increase the T_m by 10°C to 15°C. RNA:RNA hybrids increase the T_m by 20°C to 25°C. The melting temperature will also be different if PNA or other alternative nucleic acids are used as probes.[13]

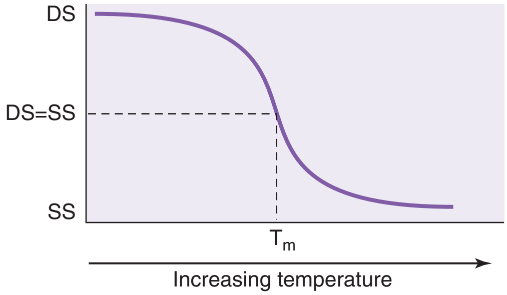


FIGURE 5.14 Melting temperature, T_m, is the point at which exactly half of a double-stranded sequence becomes single stranded. The melting temperature is determined at the inflection point of the melt curve. *DS,* double stranded; *SS,* single stranded.

The T_m is also a function of the extent of complementarity between the sequence of the probe and that of the target sequence. For each 1% difference in sequence, the T_m decreases 1.5°C.

As the length of probes decreases, the sequence becomes more influential for the final T_m. The T_m for short probes (14 to 20 bases) can be estimated by a simpler formula:

$$T_m = 4°C \times \text{number of GC pairs} + 2°C \times \text{number of AT pairs}$$

The hybridization temperature of oligonucleotide probes is about 5°C below the melting temperature.

The effect of sequence complexity on hybridization efficiency can be illustrated by the **$C_o t$ value. Sequence complexity** is the length of unique (nonrepetitive) nucleotide sequences in a genome, chromosome, or other set of double-stranded sequences. After denaturation, complex sequences require more time to reassociate than simple sequences, such as polyA:polyU. $C_o t$ is an expression of the sequence complexity (Fig. 5.15). $C_o t$ is equal to the initial DNA concentration (C_o) times the time required to reanneal it (t). **$C_o t_{1/2}$** is the time required for half of a double-stranded sequence to anneal under a given set of conditions.

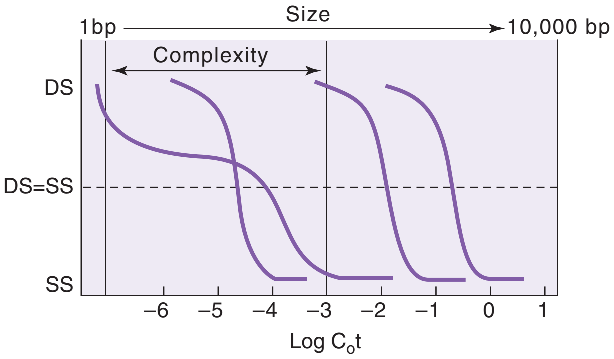


FIGURE 5.15 Reannealing of single-stranded (SS) DNA to double-stranded (DS) DNA versus time at a constant concentration yields a sigmoid curve. The complexity of the DNA sequence will widen the sigmoid curve. Increasing the length of the double-stranded DNA will shift the curve to the right.

Advanced Concepts

C_0t was used to demonstrate that mammalian DNA consisted of sequences of varying complexity. Britten and Kohne[14] used *E. coli* and calf thymus DNA to demonstrate this complexity. When they measured reassociation of *E. coli* DNA versus time, a sigmoid curve was observed, as expected for DNA molecules with equal complexity. In comparison, the calf DNA reassociation was multifaceted and spanned several orders of magnitude (see Fig. 5.15). The spread of the curve resulted from the mixture of slowly renaturing unique sequences and rapidly renaturing repeated sequences (satellite DNA), both of which are components of mammalian DNA.

T_m and C_0t values can provide a starting point for optimizing the stringency conditions for Southern or northern blot analysis. Hybridization at a temperature 25°C below the T_m for 1 to 3 $C_0t_{1/2}$ is considered optimal for a double-stranded DNA probe. The final conditions must be established empirically, especially for short probes. The stringency conditions for routine analyses, once established, are used for all subsequent assays. In the event that a component of the procedure is altered, new conditions may have to be established.

Hybridizations are generally performed in hybridization bags or in glass cylinders. Within limits, the sensitivity of the analysis increases with increased probe concentration. Because the probe is the limiting reagent, it is practical to keep the volume of the hybridization solution low. The recommended volume of hybridization buffer is approximately 10 mL/100 cm^2 of membrane surface area.

Formamide in the hybridization buffer effectively lowers the optimal hybridization temperature. This is especially useful for RNA probes and targets that, because of secondary structure, are more difficult to denature and tend to have a higher renaturation (hybridization) temperature. Incubation of the hybridization system in sealed bags in a water bath or in capped glass cylinders in rotary ovens maintains the entire membrane surface area at a constant temperature.

Short probes (less than 20 bases) can hybridize in 1 to 2 hours. In contrast, longer probes require much longer hybridization times. For Southern and northern blots with probes greater than 1,000 bases in length, incubation is carried out for 16 hours or more. Raising the probe concentration can increase the hybridization rates. Also, inert polymers, such as dextran sulfate, polyethylene glycol, or polyacrylic acid, accelerate the hybridization rates for probes longer than 250 bases.

Advanced Concepts

The nature of the probe label will affect hybridization conditions. Unlike ^{32}P labeling, the bulky nonradioactive labels (see Fig. 5.12) disturb the hybridization of the DNA chain. The temperature of hybridization with these types of probes will be lower than that used for radioactively labeled probes.

DETECTION SYSTEMS

After the transfer of gel-separated DNA, RNA, or protein to a solid membrane support and hybridization or binding of a specific probe to the target sequence of interest, the next step is to detect whether the probe has bound to the immobilized target. ^{32}P-labeled probes offered the advantages of simple and sensitive detection.

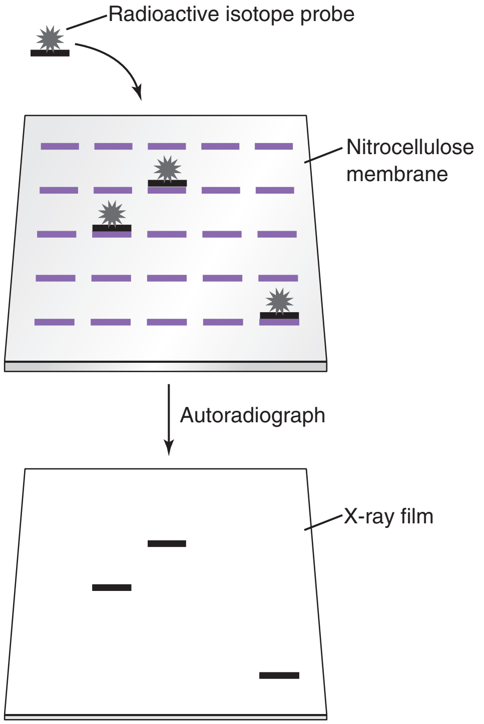


FIGURE 5.16 A DNA or RNA probe labeled with radioactive phosphorous atoms (^{32}P or ^{33}P) hybridized to target (homologous) sequences on a nitrocellulose membrane. The fragments to which the probe is bound can be detected by exposing autoradiography film to the membrane.

After hybridization, unbound probe is washed off, and the blot is exposed to light-sensitive film to detect the fragments that are hybridized to the radioactive probe (Fig. 5.16). The wash conditions must be formulated so that only completely hybridized probe remains on the blot. Typically, the wash conditions are more stringent than those used for hybridization.

Nonradioactive detection systems require a more involved detection procedure. For most nonradioactive systems, the probe is labeled with a nucleotide covalently attached to either digoxigenin or biotin. The labeled nucleotide is incorporated into the nucleotide chain of the probe by in vitro transcription, nick translation, primer extension, or addition by terminal transferase. After a digoxigenin- or biotin-labeled probe is hybridized with the blot with sample(s) containing the target sequence of interest, unbound probe is washed away. Then, anti-digoxigenin antibody or streptavidin, respectively, conjugated to alkaline phosphatase (AP conjugate; Fig. 5.17) is added to the reaction mix to bind to the digoxigenin- or biotin-labeled probe:target complex. HRP conjugates may also be used in this procedure. After the binding of the conjugate and the washing away of unbound conjugate, the membrane is bathed in a solution of substrate that, when dephosphorylated by AP or oxidized by HRP, produces a signal. Substrates frequently used are dioxetane or tetrazolium dye derivatives, which generate chemiluminescent (Fig. 5.18) or color (Fig. 5.19) signals, respectively (see Table 5.2).

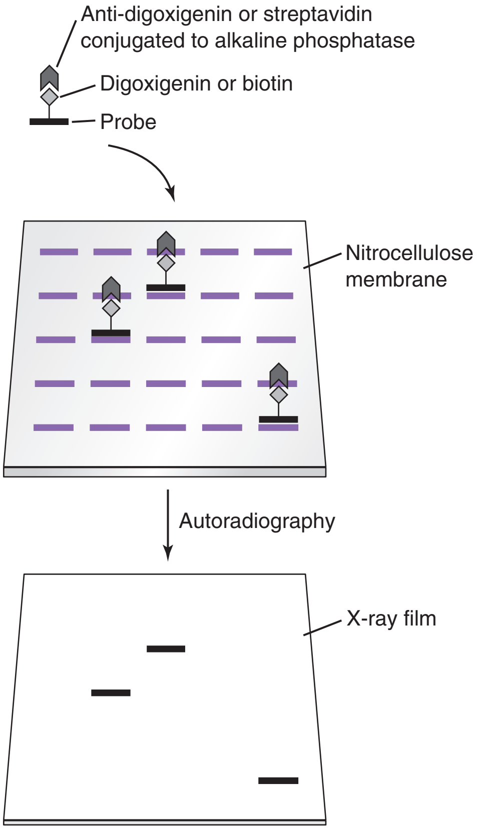


FIGURE 5.17 Indirect nonradioactive detection. The probe is covalently attached to digoxigenin or biotin. After hybridization, the probe is bound by antibodies to digoxigenin or streptavidin conjugated to alkaline phosphatase (AP). This complex is exposed to color- or light-producing substrates of AP, producing color on the membrane or light detected with autoradiography film.

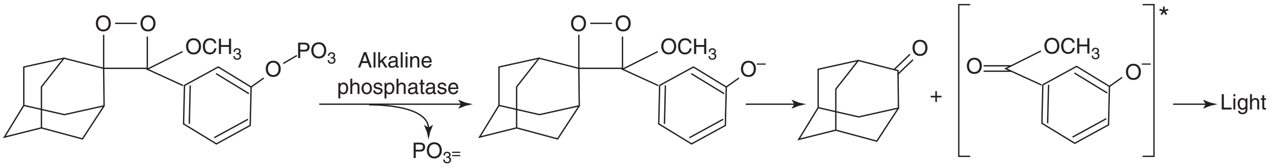

O—O
OCH3
O
PO3
Alkaline phosphatase
PO3=
O−
O
+
*
Light

FIGURE 5.18 Light is emitted from 1,2-dioxetane substrates after dephosphorylation by alkaline phosphatase to an unstable structure. This structure releases an excited anion that emits light.

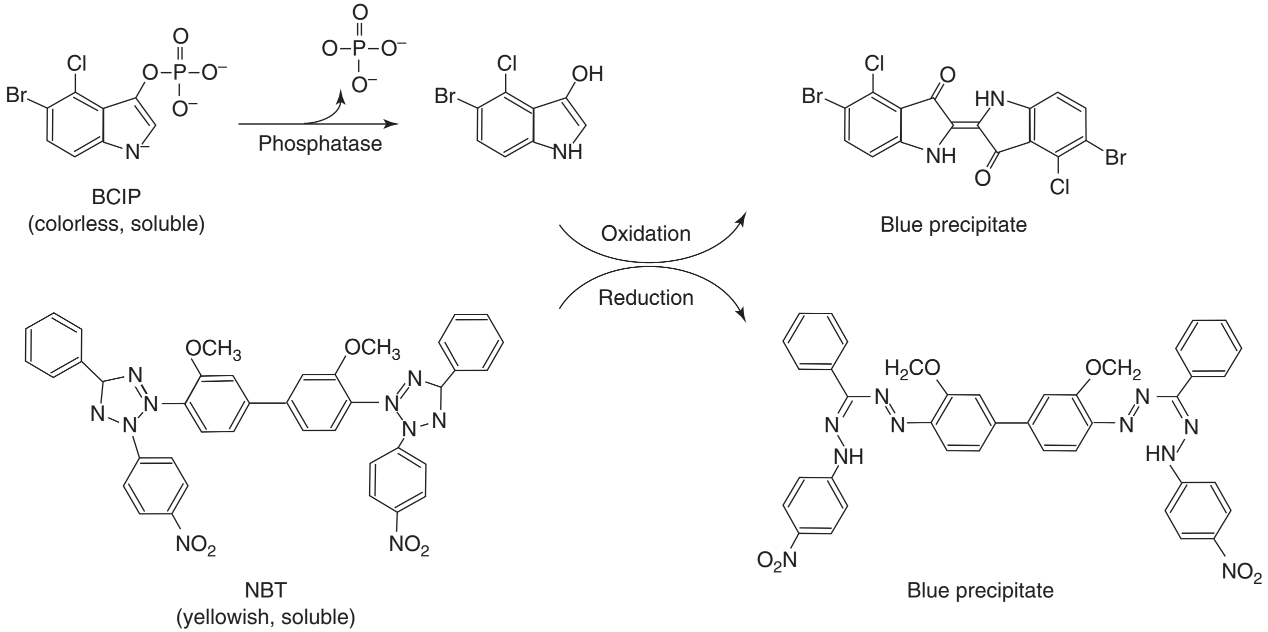

Br
Cl
O−P−O−
N−
BCIP
(colorless, soluble)
Phosphatase
OH
NH
HN
Blue precipitate
Oxidation
Reduction
OCH3
NO2
NBT
(yellowish, soluble)
H2CO
OCH2
O2N
Blue precipitate

FIGURE 5.19 Generation of color with nitroblue tetrazolium (NBT) and 5-bromo-4-chloro-3-indolyl phosphate (BCIP). Alkaline phosphatase dephosphorylates BCIP, which then reduces NBT, making an insoluble blue precipitate.

Advanced Concepts

Optimization may not completely eliminate all nonspecific binding of the probe. This will result in extra bands in control lanes due to cross-hybridizations. At a given level of stringency, any increase to eliminate cross-hybridization will lower the binding to the intended sequences. It becomes a matter of balancing the optimal probe signal with the least amount of nontarget binding. Cross-hybridizations are usually recognizable as bands of the same size in multiple runs. Cross-hybridization bands are taken into account in the final interpretation of the assay results.

Advanced Concepts

There are several substrates for chemiluminescent detection that are 1,2-dioxetane derivatives, such as 3-(2′-spiroadamantane)-4-methoxy-4-(3′-phosphoryloxy) phenyl-1,2-dioxetane (AMPPD). Dephosphorylation of these compounds by the alkaline phosphatase conjugate bound to the probe on the membrane results in a light-emitting product (see Fig. 5.18). Other luminescent molecules include acridinium ester and acridinium (*N*-sulfonyl) carboxamide labels, isoluminol, and electrochemiluminescent ruthenium trisbipyridyl labels.

TABLE 5.2 Nonradioactive Detection Systems

Type of Detection	Enzyme	Reagent	Reaction Product
Chromogenic	HRP	4-chloro-1-naphthol (4CN)	Purple precipitate
	HRP	3,3′-diaminobenzidine	Dark brown precipitate
	HRP	3,3′,5,5′,-tetramethylbenzidine	Dark purple stain
	Alkaline phosphatase	5-bromo-4-chloro-3-indolyl phosphate/nitroblue tetrazolium	Dark blue stain
Chemiluminescent	HRP	Luminol/H_2O_2/p-iodophenol	Blue light
	Alkaline phosphatase	1,2-dioxetane derivatives	Light
	Alkaline phosphatase	Disodium 3-(4-methoxyspiro {1,2-dioxetane-3,2′-(5′-chloro) tricyclo[3.3.1.1^{3,7}]decan} 4-yl)-1-phenyl phosphate and derivatives (CSPD, CDP-Star)	Light

The substrate used most often for chromogenic detection is a mixture of nitroblue tetrazolium (NBT) and 5-bromo-4-chloro-3-indolyl phosphate (BCIP). Upon dephosphorylation of BCIP by **alkaline phosphatase,** it is oxidized by NBT to give a dark blue indigo dye as an oxidation product. BCIP is reduced in the process and also yields a blue product (see Fig. 5.19).

As with radioactive detection, the chemiluminescent signal produced by the action of the enzyme on dioxetane develops in the dark by autoradiography. The release of light by phosphorylation of dioxetane occurs at the location on the membrane where the probe is bound and darkens the light-sensitive film. Chemiluminescent detection can be stronger and develops faster than radioactive detection. A disadvantage of earlier chemiluminescent detection was that it was harder to control and sometimes produced nonspecific signals. New substrates have been designed to minimize these drawbacks. Unlike radioactive detection, in which testing the membrane with a Geiger counter can give an indication of how “hot” the bands are and consequently how long to expose the membrane to the film, chemiluminescent detection may require developing films at different intervals to determine the optimum exposure time.

For chromogenic detection, color develops directly on the membrane. The advantage of this type of detection is that the color can be observed and the reaction stopped at a time when there is an optimum signal-to-background ratio. In general, chromogenic detection is not as sensitive as chemiluminescent detection and can also result in background noise, especially with probes labeled by random priming.

The key to a successful blotting method is a high signal-to-noise ratio. Ideally, the probe and detection systems should yield a specific and robust signal. A high specific signal, however, may be accompanied by background noise. Blocking agents are used with nonradioactive detection systems to avoid nonspecific binding of conjugate to the membrane; however, blocking cannot be so strong that it interferes with specific binding of the conjugate to the probe. Therefore, the specificity of detection may be sacrificed to get a stronger signal. Conversely, sensitivity may be sacrificed to generate a more specific signal.

INTERPRETATION OF RESULTS

Binding of a specific probe to its target immobilized on a membrane results in the visualization of a “band” on the membrane or film. A band is seen as a line running across the width of the lane.

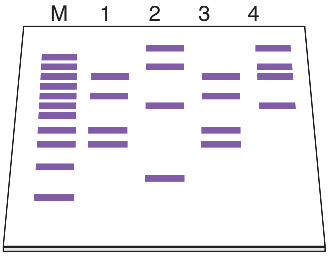


FIGURE 5.20 Example of a Southern blot result comparing restriction fragment lengths of selected regions in different samples. Restriction digests of a genetic region can also show differences in structure. The first lane (M) contains molecular-weight markers. Two samples with the same pattern (e.g., lanes 1 and 3) can be considered genetically similar.

Analysis of bands, that is, presence or absence or location in the lane, produced by Southern blot can be straightforward or complex, depending on the sample and the design of the procedure. Figure 5.20 is a depiction of a Southern blot result. The bands shown can be visualized either on a membrane or on an autoradiographic film, depending on the type of detection system. If a gene locus has a known restriction pattern, for instance, in lane 1, then samples can be tested to compare their restriction patterns. In Figure 5.20, the sample in lane 3 has the identical pattern; that is, both lanes have the same number of bands, and the bands are all in the same location on the autoradiogram and are likely to be very similar if not identical in sequence to the sample in lane 1. Southern blot cannot detect tiny deletions or insertions of nucleotides or single-nucleotide differences unless they affect a restriction enzyme site. For some assays, cross-hybridization may confuse results. These artifacts can be identified by their presence in every lane at a constant size.

Northern (or western) blots are used for the analysis of gene expression, although they can also be used to analyze transcript size, transcript processing, and protein modification. For these analyses, especially when estimating expression, it is important to include an internal control to correct for errors in isolation, gel loading, and transfer of samples. The amount of expression is then determined relative to the internal control (Fig. 5.21). In the example shown, the target transcript or protein product (top bands in all lanes, indicated by the arrow) is expressed in increasing amounts, left to right. The internal standardized control (lower bands in all lanes) ensures that a sample has low expression of target transcript or protein product and that the low signal is not due to technical difficulties.

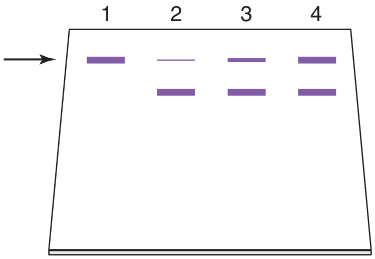


FIGURE 5.21 Example of a northern or western blot result. Lane 1 contains a positive control transcript or protein (arrow) to verify the probe specificity and target size. Molecular-weight markers can also be used to estimate size as in Southern analysis. The amount of gene product (expression level) is determined by the intensity of the signal from the test samples relative to a control gene product (lower band in lanes 2 to 4). The control transcript is used to correct for any differences in isolation or loading from sample to sample.

ARRAY-BASED HYBRIDIZATION

Dot/Slot Blots

There are many variations on hybridization configurations. In cases where the determination of the size of the target is not required, DNA and RNA can be more quickly analyzed using dot blots or slot blots. These procedures are applied to expression, mutation, and amplification/deletion analyses.

For dot or slot blots, the target DNA or RNA is deposited directly on the membrane by means of various devices, some with vacuum systems. A pipet can be used for procedures testing only a few samples. For **dot blots,** the target is deposited in a circle or dot. For **slot blots,** the target is deposited in an oblong bar (Fig. 5.22). Slot blots are more accurate for quantification by densitometry scanning because they eliminate the error that may arise from scanning through a circular target. Dot blots are useful for multiple qualitative analyses where many targets are being compared, such as in mutational screening.

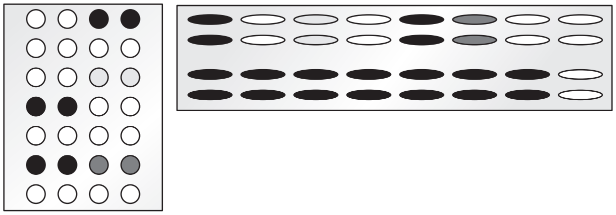


FIGURE 5.22 Example configuration of a dot blot *(left)* and a slot blot *(right)*. The target is spotted in duplicate, side by side, on the dot blot. The last two rows of spots contain positive, sensitivity, and negative control followed by a blank with no target. The top two rows of the slot blot gel on the left represent four samples spotted in duplicate, with positive, sensitivity, and negative control followed by a blank with no target in the last four samples on the right. The bottom two rows represent a loading or normalization control that is often useful in expression studies to confirm that equal amounts of DNA or RNA were spotted for each test sample.

Dot and slot blots are performed most efficiently on less complex samples, such as cloned plasmids, PCR products, or selected mRNA preparations. Without gel resolution of the target fragments, it is important that the probe hybridization conditions be optimized because cross-hybridizations cannot be definitively distinguished from true target identification. A negative control (DNA of equal complexity but without the targeted sequence) serves as the baseline for interpretation of these assays. When performing expression analysis by slot or dot blots, it is also important to include an amplification or normalization control, as shown on the slot blot in Figure 5.22. This allows correction for loading or sample differences. This control can also be analyzed on a separate, duplicate membrane to avoid cross-reactions between the test and control probes.

Genomic Array Technology

Array technology is a type of hybridization analysis allowing the simultaneous study of large numbers of targets (or samples). Arrays are applied to gene (DNA) amplification or deletion on comparative genome hybridization arrays and to gene-expression (RNA or protein) analysis on expression arrays. There are several approaches to array technology, including **macroarrays,** microarrays, high-density oligonucleotide arrays, and **microelectronic arrays.**

Macroarrays

In contrast to northern and Southern blots, dot (and slot) blots offer the ability to test and analyze larger numbers of samples at the same time. These methodologies are limited, however, by the area of the substrate material, the nitrocellulose membranes, and the volume of hybridization solution required to provide enough probe to produce an adequate signal for interpretation. In addition, although up to several hundred test samples can be analyzed simultaneously on a dot blot, those samples can be tested for only one gene or gene product.

A variation of this technique is the **reverse dot blot,** in which many different unlabeled probes are immobilized on the membrane, and the test sample is labeled for hybridization with the immobilized probes. In this configuration, the terminology can be confusing. The immobilized probe is now effectively the target, and the labeled specimen DNA, RNA, or protein is actually the probe(s). Regardless of the designation, the general idea is that a known sequence is immobilized at a known location on the blot, and the amount of sample that hybridizes to it is determined by the signal from the labeled sample.

Macroarrays are reverse dot blots of up to several thousand targets on nitrocellulose membranes. Radioactive or chemiluminescent signals were typically used to detect hybridization of a labeled sample to the target probes on the membrane. Macroarrays were created by spotting multiple probes onto nitrocellulose membranes. The hybridization of labeled sample material was read by eye or with a phosphorimager (a quantitative imaging device that uses storage **phosphor** technology instead of x-ray film). For analysis, the signal intensity from test "spots" was compared to control samples spotted on duplicate membranes.

Although macroarrays greatly increased the capacity to assess numerous targets, this system was limited by the area of the membrane and the specimen requirements. As the target number increased, the volume of sample material required increased. This limits the utility of this method for small amounts of test material, especially as might be encountered with clinical specimens.

Microarrays

In 1987, treated glass replaced nitrocellulose or nylon membranes for the production of arrays, increasing the

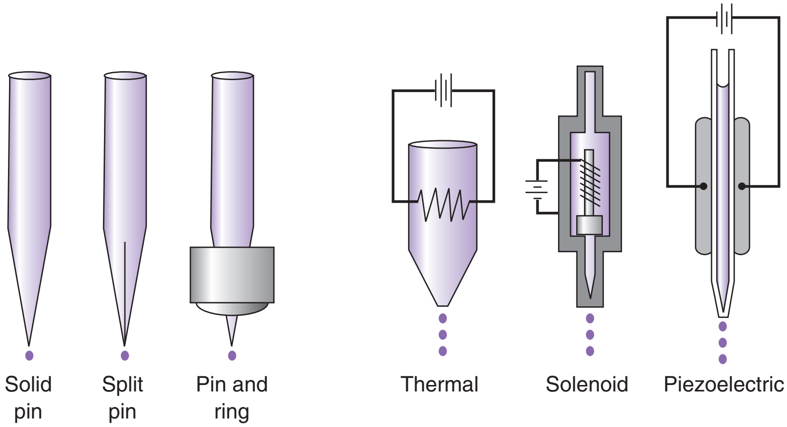


FIGURE 5.23 Pen-type *(left)* and ink-jet *(right)* technologies used to spot arrays. Pen-type spotters physically contact the array surface, in contrast to the ink-jet spotters, which do not.

versatility of array production. With improved spotting technology and the ability to deposit very small target spots on glass substrates, the macroarray evolved into the **microarray.** Tens of thousands of targets could be screened simultaneously in a very small area by miniaturizing the deposition of droplets (Fig. 5.23). Automated depositing systems (arrayers) can place more than 80,000 spots on a glass substrate the size of a microscope slide. The completion of the rough draft of the human genome sequence revealed that the human genome may consist of fewer than 30,000 genes. Thus, even with spotting representative sequences of each gene in triplicate, simultaneous screening of the entire human genome on a single chip was within the scope of array technology.

Advanced Concepts

The first automated arrayer was described in 1995 by Patrick Brown at Stanford University.[15] This and later versions of automated arrayers used pen-type contact to place a dot of probe material onto the substrate. Modifications of this technology include the incorporation of ink-jet printing systems to deposit specific targets at designated positions using thermal, solenoid, or piezoelectric expulsion of the target material (see Fig. 5.23).

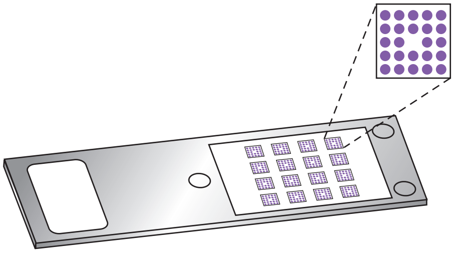


FIGURE 5.24 A microarray, or DNA chip, is a glass slide carrying hundreds to thousands of probes. Arrays are sometimes supplied with fluorescent nucleotides for use in labeling the test samples and software for identification of the probes bound with sample on the array by the array reader.

The larger nitrocellulose membrane of the macroarray, then, was replaced by a glass microscope slide of the microarray. The slide carrying the array of targets was sometimes referred to as a chip (Fig. 5.24). This terminology has led to some confusion of microarray technology with lab-on-a-chip technology. Lab-on-a-chip is a system of channels and reservoirs etched into glass or other material where chemical reactions can take place. It has no relationship to microarray chips in this regard.

Array targets immobilized on the glass slide are usually DNA—either cDNAs, PCR products, or oligomers—however, targets can be DNA, RNA, or protein. Targets are spotted in triplicate and spaced across the chip to avoid any geographic artifacts that may occur from uneven hybridization or other technical problems.

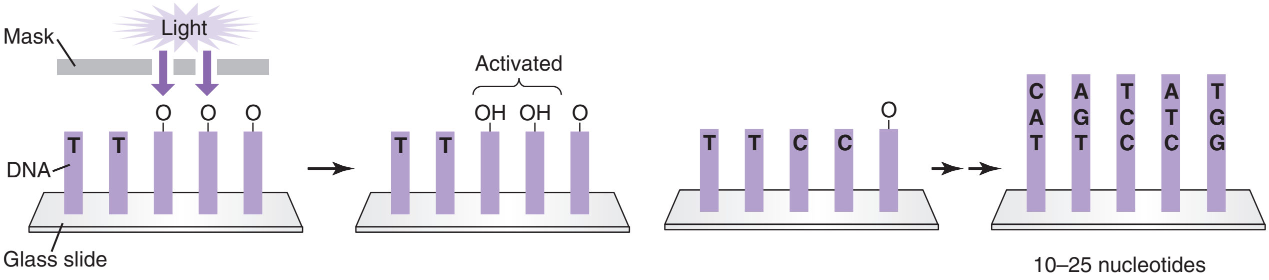


FIGURE 5.25 Photolithographic target synthesis. A mask *(left)* allows light activation of the chip. When a nucleotide is added, only the activated spots will covalently attach it *(center right)*. The masking/activation process is repeated until the desired sequences are generated at each position on the chip *(right)*.

Test samples are usually cDNA-generated from sample RNA but can, as well, be genomic DNA, RNA, or protein.

Advanced Concepts

The study of the entire genome or sets of related genes is the field of genomics. When the combinatorial and interrelated functions of gene products are known, observing the behavior of sets of genes or whole genomes is a more accurate method for analyzing biological states or responses than isolated studies of single genes. The complete set of transcripts encoded by a genome is the transcriptome, and the set of encoded proteins is the **proteome.** One gene can encode multiple transcripts so that the transcriptome is more complex than the genome. One transcript can give rise to more than one protein, making the proteome even more complex. Stanley Fields[16] predicted that the proteome is likely to be 10 times more complex than the genome. The study of entire sets of proteins, or **proteomics,** is facilitated by array technology using antigen/antibody or receptor/ligand binding in the array format and mass spectrometry.

Another method used to deposit targets for array analysis is DNA synthesis directly on the glass or silicon support.[17] This technique uses sequence information to design oligonucleotides and to selectively mask, activate, and covalently attach nucleotides at designated positions on the chip. Proprietary photolithography techniques allow for the highly efficient synthesis of short oligomers (10 to 25 bases long) on high-density arrays (Fig. 5.25). These oligomers are then probed with labeled fragments of the test sequences. Using this technology, hundreds of thousands of targets can be applied to chips with high (single-base-pair) resolution. These types of **high-density oligonucleotide arrays** are used for mutation and polymorphism analysis, DNA methylation analysis, and sequencing.

Sample preparation for array analysis requires fluorescent labeling of the test sample because microarrays and other high-density arrays are read by automated fluorescent detection systems. The most frequent labeling method used for RNA is the synthesis of cDNA or RNA copies with the incorporation of labeled nucleotides. For DNA, random priming or nick translation is used. Several alternative methods have also been developed.[18,19]

For gene-expression analyses, target probes immobilized on the chips are hybridized with labeled mRNA (cDNA) from treated cells or different cell types to assess the expression activity of the genes represented on the chip. Arrays used for this application are classified as **expression arrays**.[20] Expression arrays measure transcript or protein production relative to a reference control for each target gene isolated from untreated or normal specimens (Fig. 5.26). For proteins, immobilized antibodies (or protein ligands) are the probes, similar to their use in enzyme-linked immunoassays and western blots.

In contrast, **comparative genome hybridization** (array CGH) is designed to test DNA. This method is used to screen the genome or specific genomic loci for

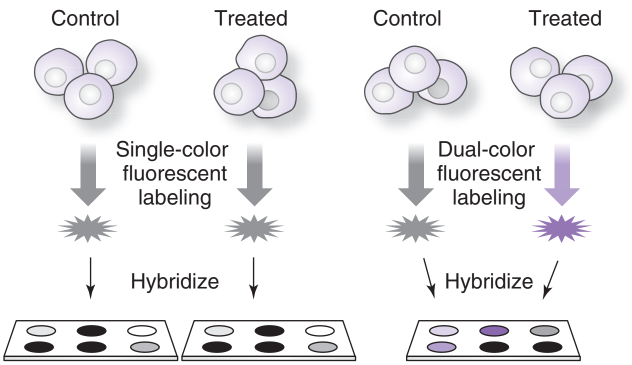

Control
Treated
Control
Treated
Single-color fluorescent labeling
Dual-color fluorescent labeling
Hybridize
Hybridize

FIGURE 5.26 Samples are labeled, rather than probes, for array analysis. At the left is single-color fluorescent labeling, where duplicate chips are hybridized separately and compared. On the right is dual-color labeling, where test (treated) and reference (control) samples are labeled with different color fluors and hybridized to the same chip.

deletions and amplifications.[21] In this method, genomic DNA is isolated, fragmented, and labeled for hybridization on the chip (Fig. 5.27), which is analogous to the cytogenetic technique done on metaphase chromosomes. Array CGH can provide higher resolution and more defined genetic information than traditional cytogenetic analysis, but it is limited to the analysis of loci represented on the array. An advantage for clinical applications is that genomic arrays can be performed on fixed tissue and limiting samples. Methods have been developed to globally amplify whole genomes to enhance CGH analysis for limited samples, such as cell-free DNA and circulating tumor cells.[22,23]

Reading microarrays requires a fluorescent reader and analysis software. After correction for background noise and normalization with standards, the software averages signal intensity from duplicate or triplicate sample data. The results are reported as a relative amount of the reference and test signals. Depending on the program, variances of more than 2 to 3 standard deviations from 1 (test = reference) are considered an indication of significant increases (test/reference greater than 1) or decreases (test/reference less than 1) in the test sample.

Several limitations of the array technology initially restrained the use of microarrays in the clinical laboratory. The lack of established standards and controls for optimal binding prevented the calibration of arrays from one laboratory to another, and not enough data had been accumulated to determine the degree of nonspecific binding and cross-hybridization that might occur among and within a given set of sequences on an array. For instance, how much variation would result from comparing two normal samples together multiple times? Background noise also affected the interpretation of array results. Furthermore, as a result of passive hybridization, different sequences will have different binding affinities under the same stringency conditions unless immobilized sequences are carefully designed to have similar melting temperatures. For mutation analysis, the length of the immobilized probe is limited due to the use of a single hybridization condition for all sequences. For gene-expression applications, only relative, rather than absolute, quantification is possible.

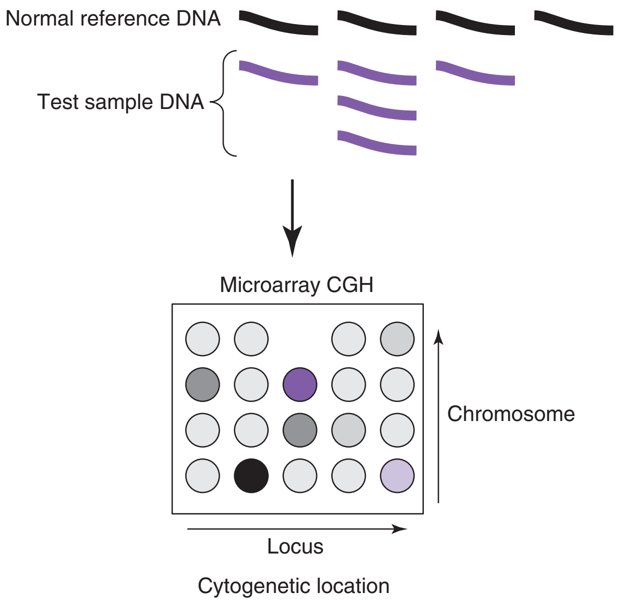


FIGURE 5.27 Comparative genomic hybridization. Reference and test DNA are labeled with different fluors, represented here as black and purple, respectively. After hybridization, excess purple label indicates amplification of the test sample locus. Excess black label indicates deletion of the test sample locus. Neutral or gray indicates equal test and reference DNA.

These and other concerns have been addressed to improve the reliability and consistency of array analysis. Baseline measurements, universal standards, and recommended controls have been established. Arrays are seeing increased use in clinical laboratories due to improvements in price, applications, and availability of instrumentation. Premade chips increase opportunities for medical applications, from targeted pathway analysis to genome-wide studies. Minimal sample requirements and comprehensive analysis with relatively small investments in time and labor are attractive features of array technology.

Bead Array Technology

Analysis of multiple targets in a single specimen is the key advantage of array technology. In microarrays, the probes are immobilized on a solid support. The probes may also be immobilized on beads, allowing hybridization of the targets in the bead suspension. In this way, multiple suspensions can be tested simultaneously, for example, in each well of a 96-well plate. In order to distinguish specific probes carried on different beads, the beads are color-coded with a particular shade of red fluorescent dye. The sample is then labeled with a green dye so that the combination of the target and bead fluorescent signals indicates the presence or absence of a specific target. This technology is used for protein and nucleic acid targets. Clinical tests using bead array systems are available for infectious diseases and tissue typing.

SOLUTION HYBRIDIZATION

In **solution hybridization,** neither the probe nor the target is immobilized. Probes and targets bind in solution, followed by resolution of the bound products.

Solution hybridization has been used to measure mRNA expression, especially when there are low amounts of target RNA. One version of the method is called **RNase protection,** or S1 mapping, for the S1 single-strand–specific nuclease. In S1 mapping, the labeled probe is hybridized to the target sample in solution. After digestion of excess probe by a single-strand–specific nuclease, the resulting labeled, double-stranded fragments are resolved by polyacrylamide gel electrophoresis (Fig. 5.28). S1 mapping is useful for determining the start point or termination point of transcripts. Nuclease protection assays are also used to detect and quantify specific RNA targets from complex RNA mixtures. This technique is now performed using commercial reagent sets.[24] The procedure is more sensitive than northern blotting because no target can be lost during electrophoresis and blotting. It is more applicable to RNA expression analysis due to limited sensitivity with double-stranded DNA targets.

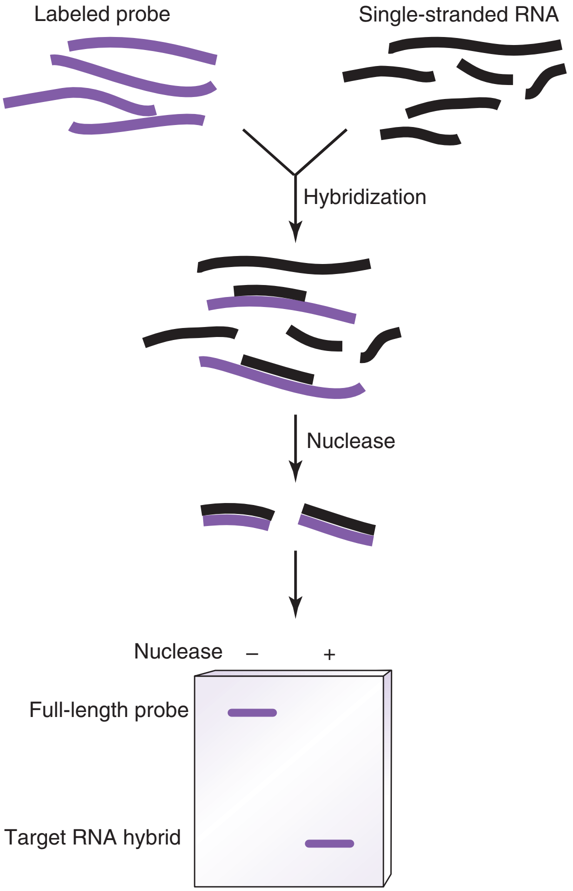


FIGURE 5.28 Solution hybridization. Target RNAs are hybridized to a labeled RNA or DNA carrying the complementary sequence to the target. After digestion by a single-strand-specific nuclease, only the target:probe double-stranded hybrid remains. The hybrid can be visualized by the label on the probe after electrophoresis.

Another variation of solution hybridization is the capture of DNA probe:RNA target hybrids on a solid support or beads rather than by electrophoresis.[25] For these "sandwich"-type assays, two probes are used, both of which hybridize to the target RNA. One probe, the capture probe, is biotinylated and will bind specifically to streptavidin immobilized on a plate or on magnetic

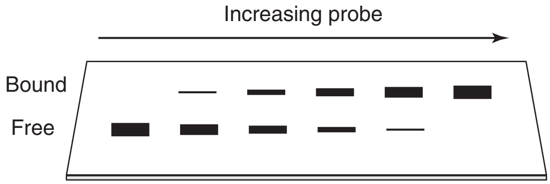


FIGURE 5.29 Gel mobility shift assay showing protein–protein or protein–DNA interaction. The labeled test substrate is mixed with the probe in solution and then analyzed on a polyacrylamide gel. If the test protein binds the probe protein or DNA, the protein will shift up in the gel assay.

beads. The other probe, called the detection probe, is detected by a monoclonal antibody directed against RNA:DNA hybrids or a covalently attached digoxigenin molecule used to generate a chromogenic or chemiluminescent signal.

Solution hybridization has also been applied to the analysis of protein–protein interactions and to nucleic acid–binding proteins using a **gel mobility shift assay.**[26,27] After mixing the labeled DNA or protein with the test material, such as a cell lysate, a change in **mobility,** usually a shift to slower migration, indicates binding of a component in the test material to the probe protein or nucleic acid (Fig. 5.29). This assay has been used to identify trans factors that bind to cis-acting elements that control gene regulation. Solution hybridization can also be used to detect sequence changes in DNA or mutational analysis. Hybridization methods offer the advantage of direct analysis of nucleic acids at the sequence level without cloning of target sequences. The significance of hybridization methodology to clinical applications is the direct discovery of molecular genetic information from routine specimen types. Widely varying modifications of the basic blotting methods have been and will be developed for clinical and research applications. Although amplification methods, specifically the PCR, have replaced many blotting procedures, some hybridization methods are still used in routine clinical analysis.

STUDY QUESTIONS

1. Calculate the melting temperature of the following DNA fragments using the sequences only:
   a. AGTCTGGGACGGCGCGGCAATCGCA
   TCAGACCCTGCCGCGCCGTTAGCGT
   b. TCAAAAATCGAATATTTGCTTATCTA
   AGTTTTTAGCTTATAAACGAATAGAT
   c. AGCTAAGCATCGAATTGGCCATCGTGTG
   TCGATTCGTAGCTTAACCGGTAGCACAC
   d. CATCGCGATCTGCAATTACGACGATAA
   GTAGCGCTAGACGTTAATGCTGCTATT

2. What is the purpose of denaturation of a double-stranded target DNA after electrophoresis and prior to transfer in a Southern blot?

3. Name two ways to permanently bind nucleic acid to nitrocellulose following transfer.

4. If a probe for a Southern blot is dissolved in a hybridization buffer that contains 50% formamide, is the stringency of hybridization higher or lower than if there was no formamide?

5. If a high concentration of NaCl was added to a hybridization solution, how would the stringency be affected?

6. Does an increase in temperature from 65°C to 75°C during hybridization raise or lower the stringency?

7. At the end of the Southern blot procedure, what would the autoradiogram show if the stringency was too high?

8. A northern blot is performed on an RNA transcript with the sequence GUAGGUAUGUA UUUGGGCGCGAACGCAAAA. The probe sequence is GUAGGUAUGUAUUUGGGCGCG. Will this probe hybridize to the target transcript?

9. In an array CGH experiment, three test samples were hybridized to three microarray chips. Each chip was spotted with eight gene probes (Genes *A–H*). The following table shows the results of this assay expressed as the ratio of test DNA to reference DNA. Are any of the eight genes

consistently deleted or amplified in the test samples? If so, which ones?

Gene	Sample 1	Sample 2	Sample 3
A	1.06	0.99	1.01
B	0.45	0.55	0.43
C	1.01	1.05	1.06
D	0.98	1.00	0.97
E	1.55	1.47	1.62
F	0.98	1.06	1.01
G	1.00	0.99	0.99
H	1.08	1.09	0.90

10. What are two differences between CGH arrays and expression arrays?

References

1. Wei H, Therrien C, Blanchard A, Guan S, Zhu Z. The Fidelity Index provides a systematic quantitation of star activity of DNA restriction endonucleases. *Nucleic Acids Research* 2008;36: e50.
2. Carte J, Christopher RT, Smith JT, Olson S, Barrangou R, Moineau S, Glover CV 3rd, Graveley BR, Terns RM, Terns MP. The three major types of CRISPR-Cas systems function independently in CRISPR RNA biogenesis in *Streptococcus thermophilus*. *Molecular Microbiology* 2014;93:98–112.
3. Keith J, Sande JD. TALENs: a widely applicable technology for targeted genome editing. *Nature Reviews Molecular Cell Biology* 2013;14:49–55.
4. Gilbert L, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, Stern-Ginossar N, Brandman O, Whitehead EH, Doudna JA, Lim WA, Weissman JS, Qi LS. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 2013;18:442–451.
5. Cho W, Jayanth N, Mullen S, Tan TH, Jung YJ, Cissé II. Super-resolution imaging of fluorescently labeled, endogenous RNA polymerase II in living cells with CRISPR/Cas9-mediated gene editing. *Scientific Reports* 2016;26:6.
6. Wu Y, Liang D, Wang Y, Bai M, Tang W, Bao S, Yan Z, Li D, Li J. Correction of a genetic disease in mouse via use of CRISPR-Cas9. *Cell Stem Cell* 2013;13:659–662.
7. Savic N, Schwank G. Advances in therapeutic CRISPR/Cas9 genome editing. *Translational Research* 2016;168:15–20.
8. Southern E. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* 1975;98:503–517.
9. Bowen B, Steinberg J, Laemmli UK, Weintraub H. The detection of DNA-binding proteins by protein blotting. *Nucleic Acids Research* 1980;8:1–20.
10. Doessing H, Vester B. Locked and unlocked nucleosides in functional nucleic acids. *Molecules* 2011;16:4511–4526.
11. Campbell M, Wengel J. Locked vs. unlocked nucleic acids (LNA vs. UNA): contrasting structures work towards common therapeutic goals. *Chemical Society Reviews* 2011;40:5680–5689.
12. Kohler G, Milstein C. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* 1975;256:495–497.
13. Giesen U, Kleider W, Berding C, Geiger A, Orum H, Nielsen PE. A formula for thermal stability (T^m) prediction of PNA/DNA complexes. *Nucleic Acids Research* 1998;21:5004–5006.
14. Britten RJ, Kohne DE. Repeated sequences in DNA. *Science* 1968;161:529–540.
15. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995;270:467–470.
16. Fields S. Proteomics in genomeland. *Science* 2001;291: 1221–1224.
17. Lipschultz RJ, Fodor SP, Gingeras TR, Lockhart, DJ. High density synthetic oligonucleotide arrays. *Nature Genetics* 1999;21: 20–24.
18. Maltas E, Malkondu S, Uyar P, Ozmen M. Fluorescent labelling of DNA on superparamagnetic nanoparticles by a perylene bisimide derivative for cell imaging. *Materials Science & Engineering C Materials for Biological Applications* 2014;48:86–93.
19. Hesse S, Manetto A, Cassinelli V, Fuchs J, Ma L, Raddaoui N, Houben A. Fluorescent labelling of in situ hybridisation probes through the copper-catalysed azide-alkyne cycloaddition reaction. *Chromosome Research* 2016;24:299–300.
20. Freeman W, Robertson DJ, Vrana KE. Fundamentals of DNA hybridization arrays for gene expression analysis. *BioTechniques* 2000;29:1042–1055.
21. Oostlander A, Meijer GA, Ylstra B. Microarray-based comparative genomic hybridization and its applications in human genetics. *Clinical Genetics* 2004;66:488–495.
22. Normand E, Qdaisat S, Bi W, Shaw C, Van den Veyver I, Beaudet A, Breman A. Comparison of three whole genome amplification methods for detection of genomic aberrations in single cells. *Prenatal Diagnosis* 2016;36:823–830.
23. Gasch C, Pantel K, Riethdorf S. Whole genome amplification in genomic analysis of single circulating tumor cells. *Methods in Molecular Biology* 2015;1347:221–232.
24. Eyler E. Explanatory chapter: nuclease protection assays. *Methods in Enzymology* 2013;530:89–97.
25. Rautio J, Barken KB, Lahdenperä J, Breitenstein A, Molin S, Neubauer P. Sandwich hybridisation assay for quantitative detection of yeast RNAs in crude cell lysates. *Microbial Cell Factories* 2003;2:4–13.
26. Malloy P. Electrophoretic mobility shift assays. *Methods in Molecular Biology* 2000;130:235–246.

27. Park S, Raines RT. Fluorescence gel retardation assay to detect protein-protein interactions. *Methods in Molecular Biology* 2004; 261:155–160.

28. Tanaka H, Fukuda N, Shoyama Y. Eastern blotting and immunoaffinity concentration using monoclonal antibody for ginseng saponins in the field of traditional Chinese medicines. *Journal of Agricultural and Food Chemistry* 2007;55:3783–3787.

29. Taki T, Gonzalez TV, Goto-Inoue N, Hayasaka T, Setou M. TLC blot (far-eastern blot) and its applications. *Methods in Molecular Biology* 2009;536:545–556.

30. Ishikawa D, Taki T. Thin-layer chromatography blotting using polyvinylidene difluoride membrane (far-eastern blotting) and its applications. *Methods in Enzymology* 2000;312:145–157.

Chapter 6

Nucleic Acid Amplification

Outline

- **TARGET AMPLIFICATION**
  - *Polymerase Chain Reaction*
    - Basic PCR Procedure
    - Components of PCR
    - Thermal Cyclers
    - The Reaction in PCR
    - Control of PCR Contamination
    - Mispriming
    - PCR Product Cleanup
    - PCR Modifications
  - *Transcription-Based Amplification Systems*
  - *Genomic Amplification Methods*
    - Whole-Genome Amplification
    - Emulsion PCR
    - Surface Amplification (Bridge PCR)
    - Arbitrarily Primed PCR
- **PROBE AMPLIFICATION**
  - *Ligase Chain Reaction*
  - *Strand Displacement Amplification*
  - *Qβ Replicase*
- **SIGNAL AMPLIFICATION**
  - *Branched DNA Amplification*
  - *Hybrid Capture Assays*
  - *Cleavage-Based Amplification*
  - *Cycling Probe*

Objectives

6.1 Compare and contrast the in vitro assays for amplifying nucleic acids with regard to type of target nucleic acid, principle, major elements of the procedure, type of amplicon produced, major enzyme(s) employed, and applications.

6.2 Describe examples of modifications that have been developed for PCR.

6.3 Discuss how amplicons are detected for each of the amplification methods.

6.4 Design forward and reverse primers for a PCR, given the target sequence.

6.5 Differentiate between target amplification and signal amplification.

Early analyses of nucleic acids were limited by the availability of DNA. Generating enough copies of a single gene required propagation of millions of cells in culture or isolation of large amounts of genomic DNA. If a gene had been cloned, many copies could be generated on bacterial plasmids, but this preparation was laborious, and some sequences were resistant to propagation in this manner.

The ability to amplify a specific DNA sequence opened the possibility of analyzing at the nucleotide level virtually any piece of DNA in nature. The first specific amplification method of any type was the **polymerase chain reaction (PCR)**. Other amplification methods have been developed based on making modifications to PCR. The methods that have been developed to amplify nucleic acids can be divided into three groups, based on whether the target nucleic acid itself, a probe specific for the target sequence, or the signal used to detect the target nucleic acid is amplified. These methods are discussed in this chapter.

TARGET AMPLIFICATION

Target amplification involves making many copies of a specific DNA sequence. This is analogous to growing cells in culture and allowing the cells to replicate their nucleic acid as well as themselves so that, for example, they can be visualized on an agar plate. Waiting for cells to replicate to detectable levels can take days to weeks or months, whereas replicating the nucleic acid in vitro only takes minutes to hours. PCR is the first and prototypical method for amplifying target nucleic acid.

Polymerase Chain Reaction

Kary Mullis visualized the idea of amplifying DNA in vitro in 1983 while driving one night on a California highway.[1,2] In the process of working through a mutation-detection method, Mullis realized that by adding all four nucleotides to his reaction mix, he could double his test **target,** a short region of double-stranded DNA, giving him 2^1, or 2, copies. If he repeated the process, the target would double again, giving 2^2, or 4, copies. After N doublings, he would have 2^N copies of his target. If $N = 30$ or 40, there would be millions of copies.

What Mullis had envisioned was the PCR. Over the next months in the laboratory, he synthesized short oligonucleotides (primers) complementary to sequences flanking a region of the human nerve growth factor and tried to amplify the region from human DNA, but the experiment did not work. Not sure of the nucleotide sequence information he had on the human gene, he tried a more defined target. The first successful amplification was a short fragment of the *Escherichia coli* plasmid, pBR322. The first paper describing a practical application, the amplification of beta-globin and analysis for diagnosis of patients with sickle cell anemia, was published 2 years later.[3] He called the method a "polymerase-catalyzed chain reaction" because DNA polymerase was the enzyme he used to drive the replication of DNA in a chain reaction. The name was quickly shortened to PCR. Since PCR was first performed, it has become increasingly user-friendly, more automated, and more amenable to use in a clinical laboratory.

Basic PCR Procedure

DNA replication in a cell requires an existing double-stranded DNA as the **template** to give the order of the nucleotide bases; the deoxyribonucleotide bases in the form of deoxynucleotide triphosphates (dNTPs)—dATP, dGTP, dCTP, dTTP; DNA polymerase to catalyze the addition of nucleotides to the growing **strand;** and a primer (synthesized by primase in vivo) to which DNA polymerase adds subsequent bases.

PCR essentially duplicates the in vivo replication of DNA in vitro, using the same components (Table 6.1) to replicate DNA, with the same end result: one copy of double-stranded DNA becoming two copies (Fig. 6.1). In less than 2 hours, PCR can produce millions of copies, collectively called **amplicons,** of DNA in contrast to days for a cell to produce the same number of copies in vivo. The real advantage of PCR is the ability to amplify specific targets. Just as the Southern blot first allowed analysis of specific regions in a complex background, PCR presents the opportunity to amplify and effectively clone the target sequences. The amplified target can then be subjected to innumerable analytical procedures.

To perform a PCR amplification, the components of the reaction—DNA template, short oligonucleotide primers, nucleotides, polymerase, and buffers—are

TABLE 6.1 Components of a Typical Reaction in PCR

Component	Purpose
0.25 mM each primer (oligodeoxynucleotides)	Directs DNA synthesis to the desired region
0.2 mM each dATP, dCTP, dGTP, dTTP	Building blocks that extend the primers
50 mM KCl	Monovalent cation (salt), for optimal hybridization of primers to template
10 mM Tris, pH 8.4	Buffer to maintain optimal pH for the enzyme reaction
1.5 mM MgCl^2	Divalent cation, required by the enzyme
2.5 units polymerase	The polymerase enzyme that extends the primers (adds dNTPs)
10^2–10^5 copies of template	Sample DNA that is being tested

subjected to an **amplification program,** which consists of a specified number of **cycles** that are divided into steps during which the samples are held at particular temperatures for designated times. Table 6.2 shows the steps of a common three-step PCR cycle.

The amplification program starts with one double-stranded DNA target. In the first step **(denaturation),** the double-stranded DNA is denatured into two single strands (Fig. 6.2). This is accomplished by heating the sample at 94°C to 96°C for several seconds to several minutes, depending on the template. The initial denaturation step is lengthened for genomic or other large DNA template fragments. Subsequent denaturations can be shorter.

TABLE 6.2 Elements of a PCR Cycle

Step	Temperature (°C)	Time (sec)
Denaturation	90–96	20–60
Annealing	50–70	20–90
Extension	68–75	10–60

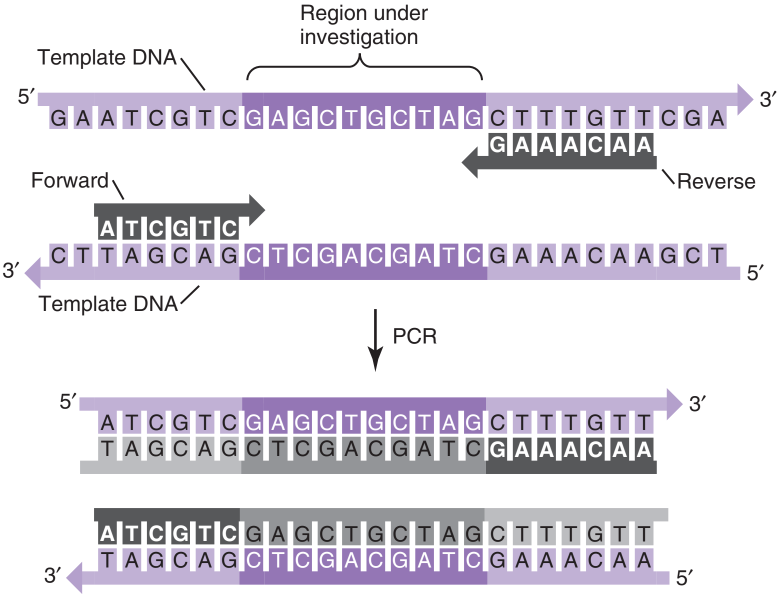

Region under investigation
Template DNA
5′
G A A T C G T C G A G C T G C T A G C T T T G T T C G A
3′
G A A A C A A
Reverse
Forward
A T C G T C
3′
C T T A G C A G C T C G A C G A T C G A A A C A A G C T
5′
Template DNA
PCR
5′
A T C G T C G A G C T G C T A G C T T T G T T
T A G C A G C T C G A C G A T C G A A A C A A
3′
A T C G T C G A G C T G C T A G C T T T G T T
T A G C A G C T C G A C G A T C G A A A C A A
3′
5′

FIGURE 6.1 The components and result of a PCR. Oligodeoxynucleotides (primers) are designed to hybridize to sequences flanking the DNA region under investigation. The polymerase extends the forward and reverse primers making many copies of the region flanked by the primer sequences, the PCR product.

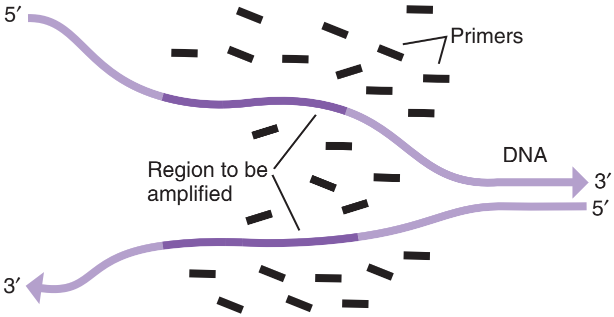


FIGURE 6.2 Denaturation of the DNA target. The region to be amplified is shown in dark purple. The primers (black) are present in vast excess.

Historical Highlights

Kary Mullis was working at Cetus Corporation, where he synthesized short, single-stranded DNA molecules, or oligodeoxynucleotides (oligos), used by other laboratories. Mullis also tinkered with the oligos he made. One night, as he drove through the mountains of northern California, Mullis thought about a method he had designed to detect mutations in DNA. His scheme was to add radioactive dideoxynucleotides—ddATP, ddCTP, ddGTP, ddTTP—to four separate DNA synthesis reactions containing oligos, the test DNA template, and polymerase. In each reaction, the complementary oligo would hybridize to the template, and the polymerase would extend the oligo with the dideoxynucleotide—but only the dideoxynucleotide that was complementary to the next nucleotide in the template. He could then determine in which of the four tubes the oligo was extended with a radioactive ddNTP by gel electrophoresis.

He thought he might improve the method by using a double-stranded template and priming synthesis on both strands, instead of one at a time. Because the results of the synthesis reaction would be affected by contaminating deoxynucleotides (dNTPs) in the reagent mix, Mullis considered running a preliminary reaction without the ddNTPs to use up any dNTPs present. He would then heat the reaction to denature the dNTP-extended oligos and add an excess of unextended oligos and the ddNTPs. As he further considered the modification to his method, he realized that if the extension of an oligo in the preliminary reaction crossed the point where the other oligo was bound on the opposite strand, he would make a new copy of the region between the oligos. He considered the new copy an additional advantage because it would improve the sensitivity of this method by doubling the target. Then, he thought, what if he did it again? The target would double again. If he added dNTPs intentionally, he could do it over and over again, doubling each time and making 2^N copies, where N is the number of times the region is replicated.

In his own words: “I stopped the car at mile marker 46,7 on Highway 128. In the glove compartment I found some paper and a pen. I confirmed that two to the tenth power was about a thousand and that two to the twentieth power was about a million and that two to the thirtieth power was around a billion, close to the number of base pairs in the human genome. Once I had cycled this reaction thirty times I would be able to reproduce the sequence of a sample with an immense signal and almost no background.”[2]

The next step in the amplification program, most critical for the specificity of the PCR, is the **annealing** step. This is the second step of the PCR cycle where the two oligonucleotide primers that will prime the synthesis of DNA anneal (hybridize) to complementary sequences on the template (Fig. 6.3). The primers dictate the part of the template that will be amplified; in other words, the primers determine the specificity of the amplification. It is important that the annealing temperature be optimized with the primers and reaction conditions. Annealing temperatures will range from 50°C to 70°C and are usually established empirically. A starting point can be determined using the T_m of the primer sequences. Reaction conditions, salt concentration, primer sequence mismatches, template condition, and secondary structure will all affect the real T_m of the primers in the reaction.

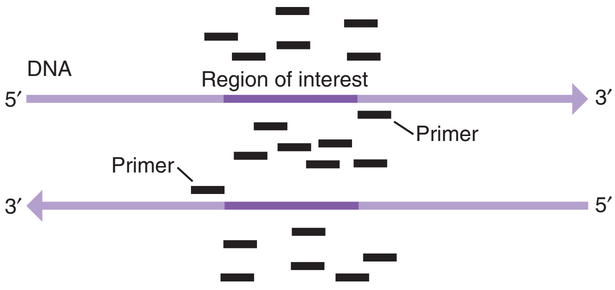


FIGURE 6.3 In the second step of the PCR cycle, annealing, the primers hybridize to their complementary sequences on each strand of the denatured template. The primers are designed to hybridize to the sequences flanking the region of interest.

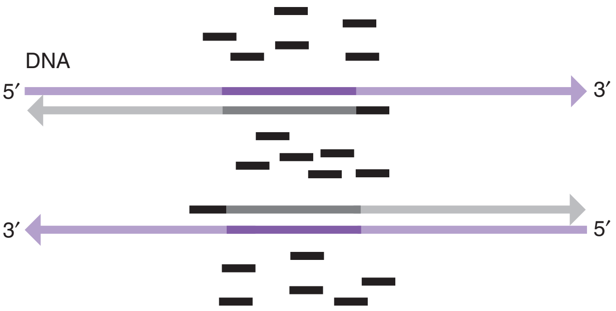


FIGURE 6.4 DNA polymerase catalyzes the addition of deoxynucleotide triphosphates (dNTPs) to the primers, using the sample DNA as the template. This completes one PCR cycle. In the original template, there was one copy of the target region. Now, after one cycle, there are two copies.

Historical Highlights

Mullis's original method, using ddNTPs and oligos to detect mutations, is still in use today. For example, fluorescent polarization–template-directed dye terminator incorporation uses fluorescently labeled ddNTPs to distinguish which ddNTP is added to the oligo. Another extension/termination assay, Homogeneous MassExtend, is a similar method, using mass spectrometry to analyze the extension products. Both of these methods were used in the Human Haplotype Mapping (HapMap) Project, which mapped millions of single-nucleotide differences in human DNA populations.

The third and last step of the PCR cycle is the primer-extension step (Fig. 6.4). This is where DNA synthesis occurs. In this step, the polymerase synthesizes a copy of the template DNA by adding nucleotides to the hybridized primers. DNA polymerase replicates the template DNA by simultaneously extending the primers on both strands of the template. This step occurs at the optimal temperature of the enzyme, 68°C to 72°C.

Historical Highlights

As Kary Mullis realized early on, the key to the brilliance of PCR is that primers can be designed to target specific sequences: “I drove on down the road. In about a mile it occurred to me that the oligonucleotides could be placed at some arbitrary distance from each other, not just flanking a base pair and that I could make an arbitrarily large number of copies of any sequence I chose and what's more, most of the copies after a few cycles would be the same size. That size would be up to me. They would look like restriction fragments on a gel. I stopped the car again. Dear Thor!, I exclaimed. I had solved the most annoying problems in DNA chemistry in a single lightning bolt. Abundance and distinction. With two oligonucleotides, DNA polymerase, and the four nucleoside triphosphates I could make as much of a DNA sequence as I wanted and I could make it on a fragment of a specific size that I could distinguish easily.”[2]

At the end of the three steps, or one cycle (denaturation, primer annealing, and primer extension), one copy of double-stranded DNA has been replicated into two double-stranded copies. Returning to the denaturing temperature starts the second cycle (Fig. 6.5), with the end result being a doubling in the number of double-stranded DNA molecules again. At the end of the PCR program, millions of copies of the original region defined by the primer sequences will have been generated. Following is a more detailed discussion of each of the components of PCR.

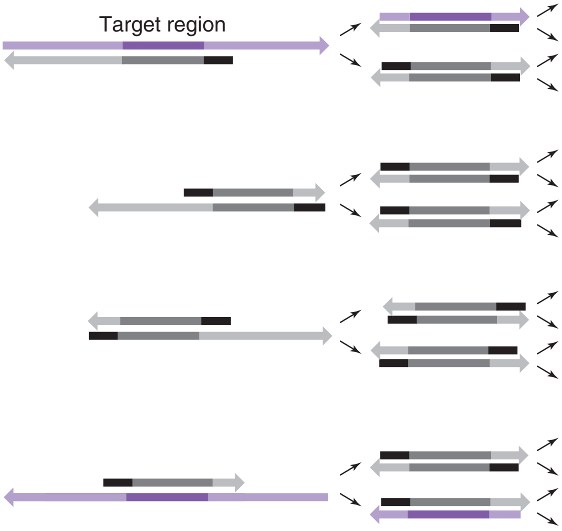


FIGURE 6.5 The products of the first cycle are again replicated in the second PCR cycle, yielding four copies of the target region. At each cycle the number of copies of the target region doubles. In an ideal PCR, the PCR product (amplicon) is composed of $2N$ copies of the target region, where N = number of PCR cycles.

Advanced Concepts

In some cases, the primer annealing temperature is close enough to the extension temperature that the reaction can proceed with only two temperature changes. This is two-step PCR, as opposed to three-step PCR that requires a different temperature for all three steps.

Components of PCR

PCR is a method of in vitro DNA synthesis. Therefore, to perform PCR, all of the components necessary for the replication of DNA in vivo are combined in optimal concentrations for replication of DNA to occur in vitro. This includes the template to be copied; primers to prime synthesis of the template; nucleotides; polymerase enzyme; and buffer components. including monovalent and divalent cations. to provide optimal conditions for accurate and efficient replication.

Primers

The primers are the critical component of the reaction because they determine the specificity of the PCR. Primers are analogous to the probes in blotting and hybridization procedures. Primers are chemically manufactured on a DNA synthesizer. Most laboratories purchase primers from commercial providers. Primer sequences are submitted in text form, along with other specifications, such as amount, degree of purification, and any modifications, such as biotin or fluorescent dyes.

Primers are designed to contain sequences complementary to sites flanking the region to be analyzed. Primer design is therefore a critical aspect of the PCR. Primers are single-stranded DNA fragments, usually 20 to 30 bases in length. The forward primer hybridizes to the complementary strand just 5′ to the sequences to be amplified. The reverse primer hybridizes just 3′ to the sequence to be amplified (see Fig. 6.1). Thus, the design of primers requires some knowledge of the target region. The placement of the primers will also dictate the size of the amplified product.

The nucleotide sequence orders of forward and reverse primers are designed using genomic sequences available from the National Center for Biological Information or other resources. When procedures are published, the primer sequences are provided in the methods descriptions. Primers can be designed manually, or alternatively, primer sequences can be automatically generated by software programs from an input target sequence and chosen parameters, such as desired amplicon length.

Hybridization of primers is subject to the same physical limitations as probe hybridization. The primer sequence (% GC) and length affect the conditions in which the primer will bind to its target. The approximate melting temperature, or T_m, of the primers can be calculated using the equation for short DNA fragments described in Chapter 5. This primer T_m then serves as a starting point for setting the optimal annealing temperature, the critical step for the specificity of the amplification reaction. Primers should be designed such that the forward and reverse primers have similar T_m so that both will hybridize optimally at the same annealing temperature. T_m can be adjusted by increasing the length of the primers or by placing the primers in areas with more or fewer Gs and Cs in the template.

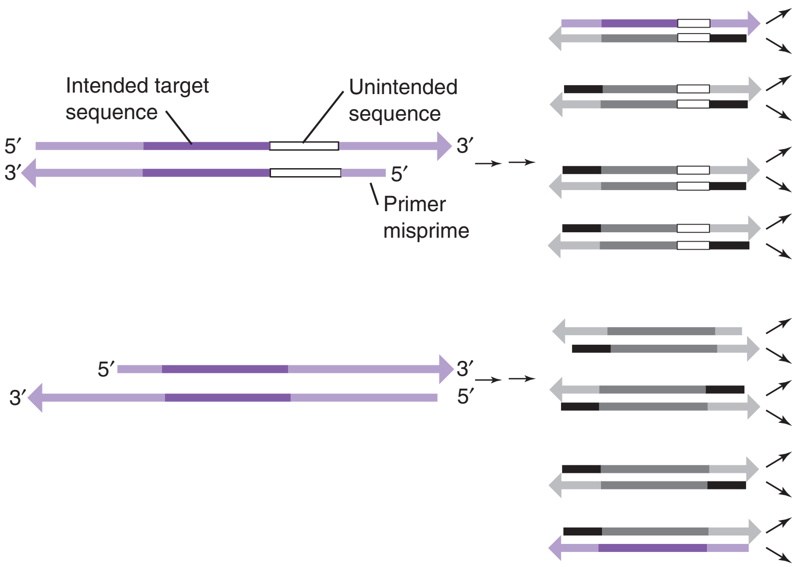


FIGURE 6.6 Mispriming of one primer creates an unintended product that could interfere with subsequent interpretation. Mispriming can also occur in regions unrelated to the intended target sequence.

Advanced Concepts

Most laboratories purchase primers from a manufacturer by submitting the required sequences, amount required (scale of synthesis), and level of purification. Standard primer orders are on the 50-to-200-nm scale of synthesis. Higher amounts (1 to 50 µM) are more expensive to purchase per base. Primers are synthesized from an immobilized 3′ end to the 5′ end so that incomplete primer molecules will be shortened from the 5′ end. Purification of the synthesized primers may be performed by cartridge or column binding and washing, high-performance liquid chromatography (HPLC), or polyacrylamide gel electrophoresis (PAGE). HPLC and PAGE purification can remove incomplete synthesis products. Primers may also be labeled at the time of synthesis with fluorescent dyes, thiolation, biotinylation, or other modifiers.

Primer sequence determines the accuracy of binding to its complementary sequence and not to other sequences. Just as cross-hybridization can occur with blot hybridization, aberrant primer binding, or **mispriming,** can occur in PCR. A fragment synthesized as a result of mispriming will carry the primer sequence and become a template for subsequent cycles of amplification (Fig. 6.6). Eventually, misprimed products will take components away from the intended reaction. They may also interfere with proper interpretation of results or with ensuing procedures, such as sequencing or mutation analysis. Secondary structure (internal folding and hybridization within DNA strands) may also interfere with PCR. Primer sequences that have internal homologies, especially at the 3′ end, or homologies with the other member of the primer pair may not work as well in the PCR. An artifact often observed in the PCR is the occurrence of **primer dimers,** which are PCR products that are approximately double the size of the primers. They result from the binding of primers onto each other through short (2- to 3-base) homologies at their 3′ ends and the copying of each primer sequence (Fig. 6.7). The resulting doublet is then a very efficient target for subsequent amplification.

The entire primer sequence does not have to bind to the template to prime synthesis; however, the 3′ nucleotide position is critical for extension of the primer. The polymerase will not form a phosphodiester bond if the 3′ end of the primer is not hydrogen-bonded to the template. This characteristic of primer binding has been exploited to modify the PCR procedure for mutation analysis of the template. There is no such strict requirement for complementarity on the 5′ end of the primer.

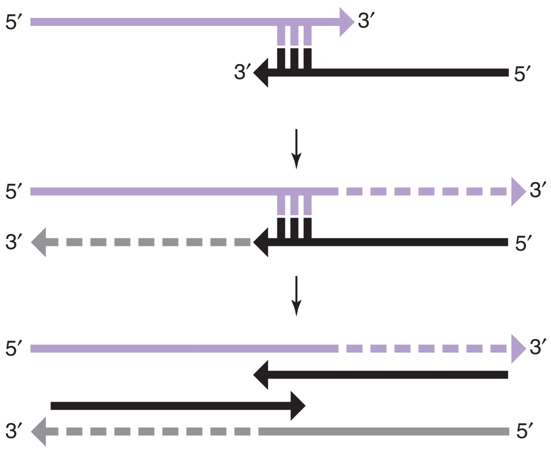


FIGURE 6.7 Formation of primer dimers occurs when there are three or more complementary bases at the 3′ end of the primers. With the primers in excess, these will hybridize during the annealing step (vertical lines), and the primers will be extended by the polymerase (dotted line) using the opposite primer as the template. The resulting product, denatured in the next cycle, will compete for primers with the intended template.

Noncomplementary extensions or tails can be added to the 5′ end of the primer sequences to introduce useful additions to the final PCR product, such as restriction enzyme sites, promoters, or binding sites for other primers. These tailed primers are designed to add or alter sequences to one or both ends of the PCR product (Fig. 6.8).

DNA Template

In a clinical sample, depending on the application, the template may be derived from the patient's genomic or mitochondrial DNA or from **viruses,** bacteria, fungi, or parasites that might be infecting the patient. Genomic DNA will have only one or two copies per cell equivalent of single-copy genes to serve as amplification targets. With robust PCR reagents and conditions, nanogram amounts of genomic DNA are sufficient for consistent results. For routine clinical analysis, 100 ng to 1 µg of DNA is usually used. Lesser amounts are required for more defined template preparations, such as cloned target DNA or product from a previous amplification.

The best templates are in good condition, free of contaminating proteins, and without nicks or breaks that can stop DNA synthesis or cause mis-incorporation of

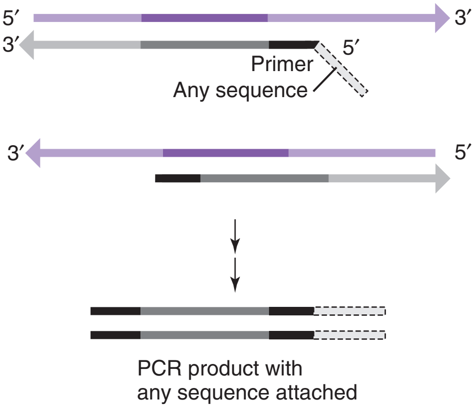


FIGURE 6.8 Sequences unrelated to the template can be added to the 5′ end of the primer. After PCR, the sequence will be on the end of the PCR product. These tailed primers can add useful sequences to one, as shown, or both ends of the PCR product. The 5′ end of the primer can also carry non-DNA molecules, such as fluorescent labels for detection of the product in capillary electrophoresis or biotin for capture of the PCR product.

nucleotide bases. Templates with high GC content and secondary structure may prove more difficult to optimize for amplification.

Advanced Concepts

Reagent systems have been designed to facilitate amplification of targets with high GC content. These systems incorporate an analog of dGTP, **deaza dGTP,** to destabilize secondary structure formed by G:C base pairing. Deaza dGTP interferes with EtBr staining in gels and is best used in procedures with other types of detection, such as autoradiography.

Deoxyribonucleotide Bases

An equimolar mixture of the four deoxynucleotide triphosphates (dNTPs)—adenine, thymine, guanine, and cytosine—is added to the synthesis reaction in concentrations sufficient to support the exponential increase of copies of the template. Standard procedures require 0.1 to 0.5 mM concentrations of each nucleotide. Substituted or labeled nucleotides such as deaza dGTP may be included in the reaction for special applications. These nucleotides will require empirical optimization for best results.

The concentration and purity of dNTP preparations affect the efficiency of the PCR reaction. A single solution containing a mixture of all four nucleotides is most convenient and lowers pipetting errors compared with four separate nucleotide solutions. The four dNTP working concentrations should be higher than the estimated K_m of each dNTP (10 to 15 mM, the concentration of substrate at half maximal enzyme velocity). Higher concentrations (at least 10 to 100 mM) are recommended for storage because storing dNTPs in lower concentrations results in hydrolysis to dNDP and dNMP. Nucleotide di- and monophosphates can also result from poor manufacturing conditions or **contamination** with heavy metals. These molecules will inhibit the PCR reaction.

Historical Highlights

The first "Molecule of the Year" selected by *Science* magazine was *Taq* DNA polymerase.[4] Development of the heat-stable enzyme is one factor responsible for the vast increase in the use of PCR from its first reported use in 1985.[3] The ability to continuously cycle through the heat steps of the PCR without loss of enzyme activity was a major advance in PCR technology.

DNA Polymerase

Automation of the PCR procedure was greatly facilitated by the discovery of the thermostable enzyme ***Taq* polymerase.** When Kary Mullis first performed PCR, he used the DNA polymerase isolated from *E. coli*. With every denaturation step, however, the high temperature inactivated the enzyme. Thus, after each round of denaturation, additional *E. coli* DNA polymerase had to be added to the tube. This was labor-intensive, lowered reaction efficiency, and provided additional opportunities for the introduction of contaminants into the reaction tube.

Taq polymerase was isolated from the thermophilic bacterium *Thermus aquaticus*. Using an enzyme derived from a thermophilic bacterium meant that the DNA polymerase could be added once at the beginning of the procedure and would maintain its activity throughout the heating and cooling cycles. Other enzymes, such as ***Tth* polymerase,** from *Thermus thermophilus*, were subsequently exploited for laboratory use. *Tth* polymerase also has **reverse-transcriptase** activity, so it can be used in **reverse-transcriptase PCR (RT-PCR)** in which the starting material is an RNA template. The addition of proofreading enzymes, for example, Vent polymerase, allows *Taq* or *Tth* polymerase to generate large products over 30,000 bases in length.

The nomenclature for the polymerases is derived from the organism from which each enzyme comes, similar to the nomenclature for restriction enzymes. For example, for the *Taq* polymerase, the *T* comes from the genus name, *Thermus*, and the *aq* comes from the species name, *aquaticus.*

Cloning of the genes coding for these polymerases has led to modified versions of the polymerase enzymes, such as the **Stoffel fragment** lacking the 289 N-terminal amino acids of *Taq* polymerase and its inherent 3′ to 5′ exonuclease activity.[5] The half-life of the Stoffel fragment at high temperatures is about twice that of *Taq* polymerase, and it has a broader range of optimal $MgCl_2$ concentrations (2 to 10 mM) than *Taq*. This enzyme is recommended for allele-specific PCR and for amplification of regions with high GC content. Further modified versions of the *Taq* enzymes retaining 3′ to 5′ exonuclease, but not 5′ to 3′ exonuclease activity, are used where high fidelity (accurate copying of the template) is important. Other variants of *Taq* polymerase, ThermoSequenase and T7 Sequenase, efficiently incorporate dideoxy nucleoside triphosphates **(NTPs)** for application to chain termination sequencing. A truncated version of *Taq* polymerase has mutations rendering it resistant to inhibitors present in whole blood, a characteristic applicable to clinical analysis.

Despite its thermal stability, the *Taq* polymerase enzyme is still subject to loss of activity under adverse conditions. For example, although mixing is important for buffers and other solutions, especially after thawing, vigorous agitation of the polymerase enzymes is not recommended. This can result in mechanical shearing, altering of the secondary and tertiary structures of complex enzymes like the polymerases, and irreversible denaturation. Mixing also introduces air to the liquid, infusing the air–liquid interface with bubbles, which can also cause damage. Most of the enzymes used in molecular biology have complex tertiary structures (with multiple subunits) and will lose enzyme activity upon vigorous mixing. In contrast, tiny proteins like RNase can renature and are therefore resistant to this type of treatment.

PCR Buffers

PCR buffers provide the optimal conditions for enzyme activity. Potassium chloride (20 to 100 mM), ammonium sulfate (15 to 30 mM), or other sources of monovalent cations are important buffer components. These salts affect the denaturing and annealing temperatures of the DNA and the enzyme activity. An increase in salt concentration makes longer DNA products denature more slowly than shorter DNA products during the amplification process, so shorter molecules will be amplified preferentially. The influence of buffer/salt conditions varies with different primers and templates.

Divalent cations, provided by magnesium chloride, also affect primer annealing and are very important for enzyme activity. Magnesium requirements will vary depending on other components of the reaction mix; for example, each NTP will take up one magnesium atom. Furthermore, the presence of ethylenediaminetetraacetic acid (EDTA) or other chelators will lower the amount of magnesium available for the enzyme. Too few Mg^{2+} ions will lower enzyme efficiency, resulting in a low yield of PCR product. Overly high Mg^{2+} concentrations promote misincorporation and thus increase the yield of nonspecific products. Lower Mg^{2+} concentrations are desirable when the fidelity of the PCR is critical. The recommended range of $MgCl_2$ concentration is 1 to 4 mM in standard reaction conditions. If the DNA samples contain EDTA or other chelators, the $MgCl_2$ concentration in the reaction mixture should be adjusted accordingly. Tris buffer and additional buffer components are also important for optimal enzyme activity and accurate amplification of the intended product; 10 mM Tris-HCl maintains the proper pH of the buffer, usually between pH 8 and pH 9.5. As with other PCR components, the optimal conditions are established empirically.

Accessory components are sometimes used to optimize reactions. Bovine serum albumin (10 to 100 μg/mL) binds inhibitors and stabilizes the enzyme. Dithiothreitol (0.01 mM) provides reducing conditions that may enhance enzyme activity. Formamide (1% to 10%) added to the reaction mixture will lower the denaturing temperature of DNA with high secondary structure, thereby increasing the availability for primer binding. Chaotropic agents (detergents), such as Triton X-100, glycerol, and dimethyl sulfoxide, added at concentrations of 1% to 10% may also reduce secondary structure to allow polymerase extension through difficult areas. These agents contribute to the stability of the enzyme as well.

Enzymes are usually supplied with buffers optimized by the manufacturer. Commercial PCR buffer enhancers of proprietary composition may also be purchased to optimize difficult reactions. Often, the buffer and its ingredients are mixed with the nucleotide bases and stored as aliquots of a master mix. The enzyme, target, and primers are then added when necessary. Dedicated master mixes will also include the primers, so only the target sequences must be added to these mixes.

Thermal Cyclers

The first PCRs were performed using multiple water baths or heat blocks set at the required temperatures for each of the steps of the PCR cycle. The tubes were manually moved from one temperature to another. In addition, before the discovery of thermostable enzymes, new enzyme had to be added after each denaturation step, further slowing the procedure and increasing the chance of error and contamination.

Automation of this tedious process was greatly facilitated by the availability of the heat-stable enzymes. To accomplish the PCR, then, an instrument must only manage temperature according to a scheduled amplification program. Thermal cyclers, or thermocyclers, were thus designed to rapidly and automatically ramp (change) to the required incubation temperatures, holding at each one for designated periods.

Early versions of thermal cyclers were designed as heater/coolers with programmable memory to record the appropriate reaction conditions. Compared with modern models, the memory available for recording the reaction conditions and sample was limited. Wax or oil (vapor barriers) had to be added to the reactions to prevent condensation of the sample on the tops of the tubes during the temperature changes. The layer of wax or oil made subsequent sample handling more difficult. Later, thermal cycler models were designed with heated lids that eliminated the requirement for vapor barriers.

The several available versions of thermal cyclers differ in heating and/or refrigeration systems as well as the programmable software within the units. Some hold samples in open chambers for air heating and cooling; others hold samples in blocks designed to accommodate 0.2-mL tubes, usually in a 96-well format. Some models have interchangeable blocks to accommodate amplification in different sizes and numbers of tubes or in situ amplification of nucleic acid targets in tissue on slides. A cycler may run more than one block independently so that different PCR programs can be performed simultaneously. Rapid PCR systems work with small sample volumes in chambers that can be heated and cooled quickly by changing the air or specially designed block temperature surrounding the samples.[6] **Real-time PCR** systems are equipped with fluorescent detectors to measure the PCR product as the reaction proceeds. PCR can also be performed in a microchip device in which 1- to 2-µL samples are forced through tiny channels etched in a glass chip, passing through temperature zones as the chip rests on a specially adapted heat block or microwave heating system.[7,8]

The Reaction in PCR

For routine PCR, an appropriate amount of DNA that has been isolated from a test specimen is mixed with the other PCR components, either separately or as a **master mix.** The final volume of the reaction mix varies from 1 to 50 µL for most PCR procedures. Thermal cyclers take thin-walled tubes, tube strips, or 96-well or 384-well plates with 0.04- to 0.2-mL volume capacity. Preparation of the specimen for PCR is classified as a pre-PCR procedure. To avoid contamination (see following discussion), it is recommended that the pre-PCR work be performed in a designated area that is clean and free of amplified products or other extraneous DNA. The sample tubes are then loaded into the thermal cycler programmed with the temperatures and times for each step of the PCR cycle, the number of cycles to be completed (usually 30 to 50), the conditions for ramping from step to step, and the temperature at which to hold the tubes once all of the cycles are complete. Because the PCR products are stable at the holding temperature, the technologist does not have to retrieve the samples from the thermal cycler immediately after the PCR program is concluded.

After PCR, a variety of methods can be used to analyze the product. Most commonly, the PCR product is analyzed by gel or capillary electrophoresis. Depending on the application, the size, presence, or intensity of PCR products is observed after electrophoresis. An example of the results from a PCR run is shown in Figure 6.9.

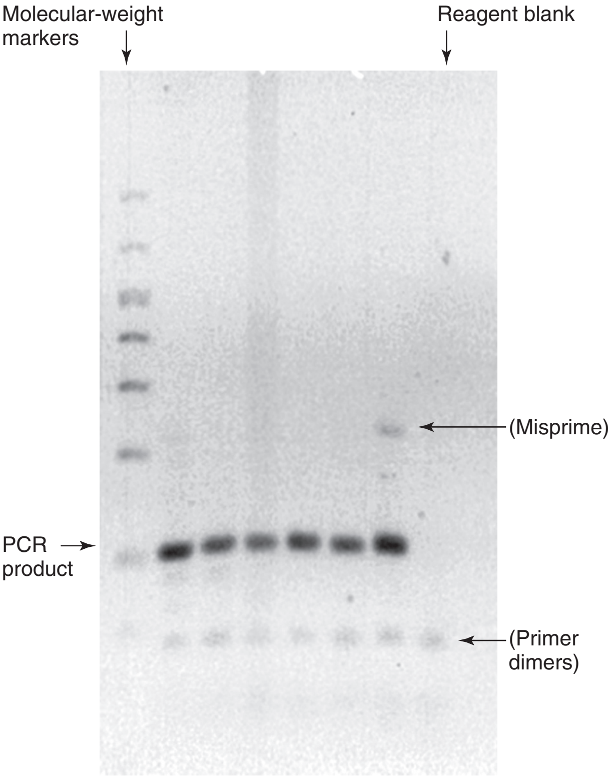


FIGURE 6.9 Example of PCR products after resolution on an agarose gel and staining with ethidium bromide. Molecular-weight markers in the first gel lane are used to estimate the size of the PCR product. The intended product is 125 bp. Artifactual primer dimers and misprimed products are also present. Absence of products in the last gel lane (reagent blank) confirms that there is no contamination in the master mix.

Controls for PCR

As with any diagnostic assay, running the correct controls during every PCR run is essential for ensuring and maintaining the accuracy of the assay. Positive controls ensure that the enzyme is active, the buffer is optimal, the primers are priming the right sequences, and the thermal cycler is cycling appropriately. A negative control without DNA (also called a **contamination control** or **reagent blank**) ensures that the reaction mix is not contaminated with template DNA or amplified products from a previous run. A negative control with DNA that lacks the target sequence **(negative template control)** ensures that the primers are not annealing to nontarget sequences of DNA. In some applications of PCR, an internal amplification control is included that contains a second set of primers and an unrelated target added to the reaction mix. Alternatively, the amplification control can be a synthetic template containing the target primer-binding sites but producing a larger product than the target analyte. These controls demonstrate that the reaction is working even if the test sample is not amplified. Amplification controls are amplified preferably in the same tube with the test reaction, although it is acceptable to perform the amplification control in a duplicate reaction mix if the control cannot easily be distinguished from the target product in the same reaction. This type of control is most important when PCR results are reported as positive or negative, "negative" meaning that the target sequences are not present. The amplification control is critical to distinguish between a true negative for the sample and an amplification failure **(false negative).**

Control of PCR Contamination

Contamination is a significant concern in "open-tube" methods that involve target amplification and manipulation of the PCR product. The nature of the amplification procedure is such that, theoretically, a single molecule will give rise to a product. This is of great concern in the medical or forensic laboratory where results may be interpreted based on the presence, absence, size, or amount of a PCR product. With modern reagent systems designed for robust amplification of challenging specimens, such as paraffin-embedded tissues or samples with low cell numbers, the balance between aggressive amplification of the intended target and avoidance of a contaminating template is delicate. For this reason, contamination control is of utmost importance in designing a PCR procedure and laboratory setup.

Although genomic DNA (e.g., from hair, skin, or ambient microorganisms) is a source of spurious PCR targets, the major cause of contamination is the presence of PCR products from previous amplifications. Unlike the relatively large and scarce genomic DNA, the small, highly concentrated PCR product DNA can aerosolize when tubes are uncapped and when the amplified DNA is pipetted. This PCR product is a perfect template for primer binding and amplification in a subsequent PCR using the same primers. Contamination control procedures, therefore, are mainly directed toward eliminating PCR product from the setup reaction.

Contamination is controlled both physically and chemically. Physically, the best way to avoid PCR carryover is to physically separate the **pre-PCR** areas from the **post-PCR** analysis areas. Positive airflow, airlocks, and more extensive measures are taken by high-throughput laboratories that process large numbers of samples and test for a limited number of amplification targets. Most laboratories can separate these areas by assigning separate rooms or using isolation cabinets. Equipment, including laboratory gowns and gloves, and reagents should be dedicated to either pre- or post-PCR. Gloves should be changed if contaminated areas are touched before proceeding with the PCR setup. Items can flow from the pre- to the post-PCR area but not in the opposite direction, without decontamination.

Ultraviolet (UV) light has been used to decontaminate and maintain pre-PCR areas.[9] UV light catalyzes single- and double-strand breaks in the DNA that will then interfere with replication. Isolation cabinets are equipped with UV light sources that are turned on for about 20 minutes after the box has been used. The effectiveness of UV light may be increased by the addition of **psoralens** to amplification products after analysis. Psoralens intercalate between the bases of double-stranded DNA, and in the presence of long-wave UV light, they covalently attach to the thymidines, uracils, and cytidines in the DNA chain. The bulky adducts of the psoralens prevent denaturation and amplification of the treated DNA.

Although convenient and effective for antimicrobial decontamination,[10] UV treatment may not be the most effective decontaminant for nucleic acid contamination.[11-13] The efficiency of UV-light treatment for decontamination depends on the wavelength, energy, and distance of the light source. The technologist must avoid skin or eye exposure to UV light, and UV light will also damage Plexiglas and some plastics, so laboratory equipment, including pipets, may be affected by extended exposure. Furthermore, the use of overhead UV room lights for decontamination is not very efficient for surface DNA decontamination due to the hazardous high intensity (4,000 microwatt-seconds/cm^2) required to damage DNA at the distances the light has to travel from the ceiling to the bench surface.

An effective method for decontamination and preparation of the workspace is 10% bleach (7 mM sodium hypochlorite). Frequently wiping bench tops, hoods, or any surface that comes in contact with specimen material with dilute bleach or alcohol removes most DNA contamination. As a practice in forensic work, before handling evidence or items that come in contact with evidence, gloves are wiped with bleach and allowed to air-dry.

An enzymatic method of contamination control is the dUTP–UNG system, which involves substitution of dUTP for dTTP in the PCR reagent master mix, resulting in incorporation of dUTP instead of dTTP into the PCR product. Although some polymerase enzymes may be more or less efficient in incorporating the nucleotide, the dUTP does not affect the PCR product in most applications. At the beginning of each PCR, the enzyme uracil-*N*-glycosylase (UNG) is added to the reaction mix. This enzyme will degrade any nucleic acid containing uracil, such as contaminating PCR product from previous reactions. A short incubation period is added to the beginning of the PCR amplification program, usually at 50°C for 2 to 15 minutes, to allow the UNG enzyme to function. The initial denaturation step in the PCR cycle will degrade the UNG before synthesis of the new products. Note that this system will not work with some types of PCR, such as nested PCR (discussed later in the chapter), because a second round of amplification requires the presence of the first-round product. The dUTP–UNG system is used routinely in quantitative PCR procedures in which contamination control is essential because the contaminant will affect the interpretation of the amount of product.

Wipe tests are performed in some laboratories as a routine check or to confirm successful decontamination. Filter paper is wiped on any exposed or touched surfaces in the pre-PCR setup, extraction, and amplification areas. The paper is then placed in robust PCR buffer with appropriate primers and subjected to amplification. Any production of amplicon indicates the presence of contamination.

Mispriming

When PCR products are analyzed for size and purity by electrophoresis, the amplicon size should agree with the size determined by the primer placement. For instance, if two 20-base primers were designed to hybridize to sequences flanking a target of 100 base pairs (bp), the amplicon should be 140 bp in size. Any larger or smaller amplicons would be due to mispriming, primer dimers, or other artifacts of the reaction. For some procedures, these artifacts do not affect the interpretation of results and can be ignored as long as they do not compromise the efficiency of the reaction. For other purposes, however, extraneous PCR products must be avoided or removed.

Misprimes are initially averted by good primer design and optimal amplification conditions. Even under the best conditions, however, misprimes can occur during preparation of the reaction mix. This is because *Taq* polymerase has some activity at room temperature. While mixes are prepared and transported to the thermal cycler, the primers and template are in contact at 22°C to 25°C, a condition of very low stringency. Under these conditions, the primers can bind sequences other than their exact complements in the target, and the low-level activity of *Taq* will extend them. These misprimed products, then, are already present before the amplification program begins. **Hot-start PCR** can be used to prevent this type of mispriming.

Advanced Concepts

In addition to breaking the sugar-phosphate backbone of DNA, UV light also stimulates covalent attachment of adjacent pyrimidines in the DNA chain, forming pyrimidine dimers. These boxy structures are the source of mutations in DNA in some diseases caused by sun exposure. DNA repair systems remove these structures in vivo. Loss of these repair systems is manifested in diseases such as xeroderma pigmentosum, Cockayne syndrome, and trichothiodystrophy.[14]

Hot-start PCR can be performed in three different ways. In one approach, the reaction mixes are prepared on ice and placed in the thermal cycler after it has been prewarmed to the denaturation temperature. A second way to perform hot-start PCR was used with older model thermal cyclers that did not have heated lids and required a vapor barrier over each reaction. A bead of wax was placed in each reaction tube containing all components of the reaction mix except enzyme and template. The tubes were heated to 100°C to melt the wax and then cooled to room temperature. The melted wax would float to the top of the reaction mix and congeal into a physical barrier as it cooled. The remainder of the reaction mix containing template and enzyme was then added on top of the wax barrier. When the tubes were placed in the thermal cycler, the wax melted at the denaturation temperature, and the primers and template first came in contact at the proper annealing temperature. The floating wax then served as an evaporation barrier as the reaction proceeded. The third and most frequently used hot-start method utilizes sequestered enzymes. These enzymes are supplied in an inactive form, sequestered and inactivated by monoclonal antibodies or by other proprietary methods. The enzyme will not extend the primers until it is activated by heat in the first denaturation step of the PCR program, thereby preventing any primer extension during the preparation of the reagent mix.

Touchdown PCR is a modification of the PCR program used to enhance the amplification of the desired PCR product. In this method, the PCR program begins with annealing temperatures higher than the optimal target primer-binding temperature. The annealing temperature is decreased by 1°C every cycle or every other cycle until the optimal annealing temperature is reached. Subsequent cycles are carried out at the optimal temperature. Any difference in the annealing temperature will translate into an exponential advantage for correct versus mismatched primer binding.[15]

PCR Product Cleanup

Sequence limitations in primer design or reaction conditions may not completely prevent primer dimers or other extraneous products. These unintended amplicons are unacceptable for analytical procedures that demand pure product, such as sequencing or certain mutation analyses. A direct way to obtain a clean PCR product is to resolve the amplification products by gel electrophoresis, cut out pieces of the gel containing desired bands, and elute the PCR product. Agarose gel slices can be digested with enzymes, such as β-agarase, or by incubation with iodine, releasing the product DNA (Fig. 6.10).

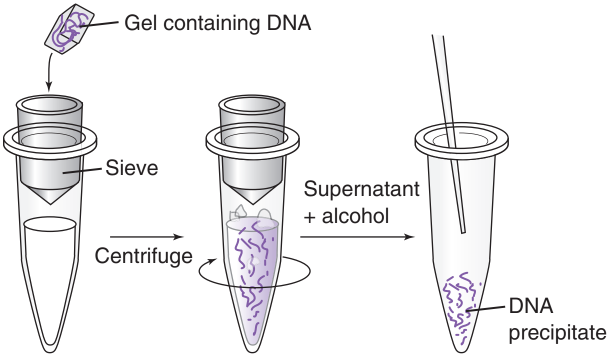


FIGURE 6.10 After gel electrophoresis, the gel band of PCR product is excised with a clean scalpel or spatula. The gel is disintegrated by centrifugation through a sieve, releasing the DNA. The DNA in solution can then be separated from the gel fragments, precipitated with alcohol, and pelleted by a second centrifugation.

Residual components of the reaction mix, such as leftover primers and unused nucleotides, also interfere with some post-PCR applications. Moreover, the buffers used for PCR may not be compatible with post-PCR procedures. Amplicons free of PCR components are conveniently prepared using **spin columns** (Fig. 6.11) or silica beads. The DNA binds to the column, and the rest of the reaction components are rinsed away by centrifugation. The DNA can then be eluted from the column. Even though columns or beads provide better recovery than gel elution, they may not completely remove residual primers or misprimed products.

Addition of alkaline phosphatase (AP) in combination with exonuclease I (ExoI) is an enzymatic method for removing nucleotides and primers from PCR products prior to sequencing or mutational analyses. During a 15- to 30-minute incubation at 37°C, AP dephosphorylates unincorporated nucleotides, and ExoI degrades the single-stranded primers. The enzymes must then be removed by extraction or inactivated by heating at 95°C for 5 minutes. This method is convenient because it is performed in the same tube as the PCR. It does not, however, remove other buffer components.

In some post-PCR methods, such a small amount of PCR product is added to the next reaction that residual components of the amplification are of no consequence, so no further cleanup of the PCR product is required. The choice of cleanup procedure or whether cleanup is necessary at all will depend on the post-PCR application.

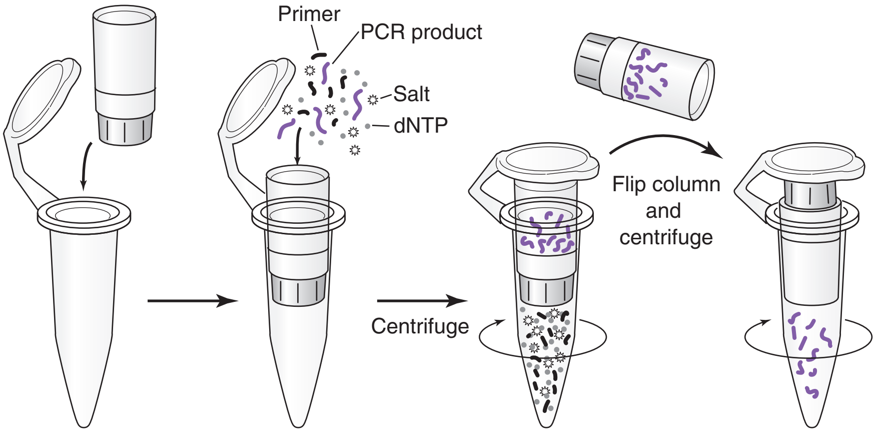


FIGURE 6.11 PCR product cleanup in spin columns *(left)* removes residual components in the PCR mix. Amplicon DNA binds to a silica matrix in the column while the buffer components flow through during centrifugation. The column is then inverted, and the DNA is eluted by another centrifugation in low salt (Tris-EDTA) buffer.

PCR Modifications

PCR has been adapted for various applications, several of which are used in the medical laboratory. Among the large and increasing number of PCR modifications are the following methods that might be encountered in the clinical molecular laboratory. These methods are capable of detecting multiple targets in a single run **(multiplex PCR)** using RNA templates (reverse transcriptase PCR) or such amplified products as templates (nested PCR) and quantifying starting template (quantitative PCR).

Multiplex PCR

More than one primer pair can be added to a PCR so that multiple amplifications are primed simultaneously, resulting in the formation of multiple products. Multiplex PCR is especially useful in typing or identification analyses. Individual organisms, from viruses to humans, can be identified or typed by observing a set of several PCR products at once. Pathogen typing and forensic identification kits contain multiple sets of primers that amplify polymorphic DNA regions. The pattern of product sizes will be specific for a given type or individual.

Multiple organisms have been the target of multiplex PCR in clinical microbiology laboratories.[16,17] One respiratory sample, for example, can be used to test for the presence of more than one respiratory virus.[18] In a slightly different approach to testing for multiple targets, one set of primers can detect an infectious organism, and a second set can then detect the presence of a gene that makes that organism resistant to a particular antimicrobial agent. Multiplex PCR reagents and conditions require more complex optimization. Target sequences may not amplify with the same efficiency, and primers may interfere with each other in binding to the target sequences. The conditions for the PCR must be adjusted for the optimal amplification of all products in the reaction. Multiplexing primers is useful not only to detect multiple targets but also to confirm accurate detection of a single target. Internal amplification controls are often multiplexed with test reactions that are interpreted by the presence or absence of product. The control primers and targets must be chosen so that they do not interfere or compete with the amplification of the test region. Internal amplification controls are the ideal for positive/negative qualitative PCR tests.

Sequence-Specific PCR

The strict requirement for complementarity of the 3′ end of primers can be used to identify single-base changes in the target DNA. By designing the forward or reverse primer to end with a 3′ base complementary to the mutant sequence, the presence of the base change is detected by successful amplification. Although this method is used frequently in the medical laboratory to detect base changes in patient DNA, its high sensitivity risks detection of mutations at very low levels that may not be clinically significant.[19] Sequence-specific PCR is one of the methods of human leukocyte antigen allele analysis in tissue typing.

Reverse Transcriptase PCR

If the starting material for a procedure is RNA, the RNA may first be converted to double-stranded DNA, which is a better template for amplification than single-stranded RNA. The conversion is accomplished through the action of reverse transcriptase (RT), an enzyme isolated from RNA viruses. This enzyme first copies the RNA single strand into an RNA:DNA hybrid and then uses a hairpin formation on the end of the newly synthesized DNA strand to prime synthesis of the complementary DNA strand, replacing the original RNA in the **hybrid.** The resulting double-stranded DNA is copy or complementary DNA (cDNA).

Like other DNA polymerases, reverse transcriptase requires priming. Gene-targeted primers—oligo dT primers or random hexamers—are most often used to prime the synthesis of the initial DNA strand. The yield of cDNA will be relatively low using gene-targeted primers but highly specific for the target of interest. The specific primers will prime cDNA synthesis only from transcripts complementary to the primer sequences. Oligo dT primers are 18-base-long single-stranded polyT sequences that will prime cDNA synthesis only from RNA with polyA tails (mRNA and some noncoding RNA). The yield of cDNA will be higher with oligo dT primers than with target-specific primers because potentially all polyA RNA could be included in the specimen. The highest yield of cDNA is achieved with random hexamers or decamers. These are 6- or 10-base-long single-stranded oligonucleotides with random sequences. The 6 or 10 bases will match and hybridize to random sites in the target RNA with some frequency, priming DNA synthesis. Random priming will generate cDNA from all RNA in the specimen.

RT PCR is used to measure RNA expression profiles, to detect rRNA, to analyze gene regions interrupted by long introns, and to detect microorganisms with RNA genomes. For gene-expression analysis, the amount of cDNA reflects the amount of transcript in the preparation. In other applications, genes that are interrupted by long introns can be made more available for consistent amplification using cDNA versions lacking the interrupting sequences. cDNA is often used for sequencing because the sequence of the coding region can be determined without long stretches of introns that might complicate the analysis. Originally, RT PCR was performed in two steps: cDNA synthesis and then PCR. *Tth* DNA polymerase, which has RT activity, and proprietary mixtures of RT and sequestered (hot-start) DNA polymerase are components of one-step RT PCR procedures.[20] These methods are more convenient than the two-step procedure because RNA is added directly to the PCR. The amplification program is modified to include an initial incubation of 45°C to 50°C for 30 to 60 minutes, during which RT makes cDNA from RNA in the sample. The RT activity will then be inactivated, and the DNA polymerase activated, in the first denaturation step of the PCR procedure.

Although RT PCR is a widely used and important adjunct to molecular analysis, it is subject to the vulnerabilities of RNA degradation. As with other procedures that target RNA, specimen handling is important for accurate results.[21] Methods have been described for the RT PCR amplification of challenging specimens, such as paraffin-embedded tissues; however, fixed specimens are difficult to analyze consistently.[22]

Nested PCR

The increased sensitivity that PCR offers is very useful in clinical applications because clinical specimens are often limited in quantity and quality. Low levels of target and the presence of interfering sequences can prevent a regular PCR from working with the reliability required for clinical applications. **Nested PCR** is a modification that increases the sensitivity and specificity of the reaction.[23-28]

In nested PCR, two pairs of primers are used to amplify a single target in two separate PCR runs. The second pair of primers, designed to bind slightly inside of the binding sites of the first pair, will amplify the product of the first PCR in a second round of amplification. The second amplification will specifically increase the amount of the intended product. In **semi-nested PCR,** one of the second-round primers is the same as the first-round primer (Fig. 6.12).

Several variations of nested and semi-nested PCR have been devised. For example, as shown in Figure 6.12, the first-round primers can have 5′ sequences added (5′ tails) complementary to sequences used for second-round primers. This tailed primer method is valuable for multiplex procedures in which multiple

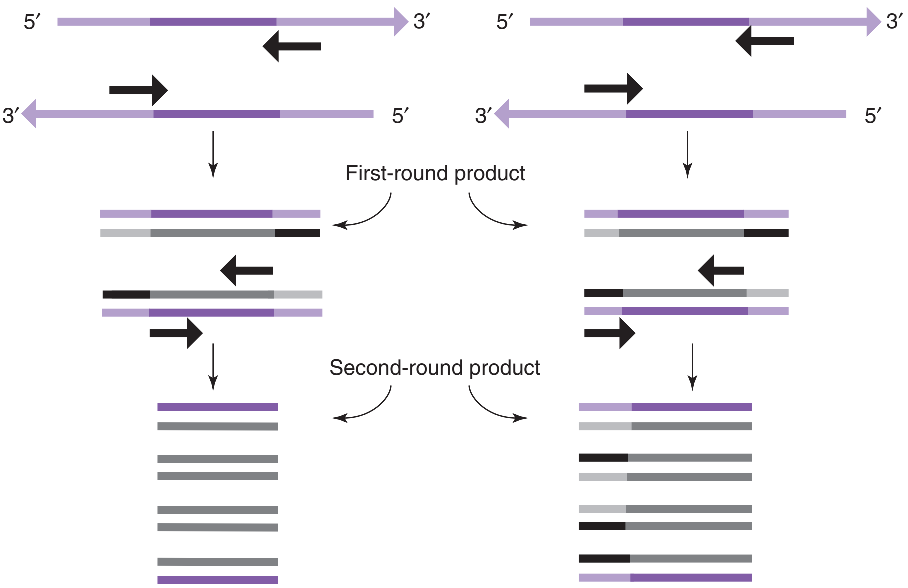


FIGURE 6.12 Variations of nested PCR using nested primers *(left),* semi-nested second-round primers *(right),* and tailed first-round primers. In nested PCR, a second reaction amplifies the product of the original reaction with primers binding sites present on the amplicon. Both second-round products are binding within the amplicon. In semi-nested PCR, one primer binds within the amplicon, and the other is identical to the first-round primer.

first-round primers may differ in their binding efficiencies. With the tailed primers, sequences complementary to a single set of second-round primers are added to all of the first-round products. In the second round, then, all products will be amplified with the same primers. Although this tailed primer procedure increases sensitivity in multiplex reactions, it does not increase specificity.

Real-Time (Quantitative) PCR

Standard PCR procedures will indicate if a particular target sequence is present in a clinical sample. For some situations, though, the clinician is also interested in how much of the target sequence is present. Initially, there were several approaches to estimating the amount of starting template via PCR. The nature of amplification, however, made calculating direct quantities of starting material complex. Strategies for quantifying starting material by quantifying the end products of PCRs utilized specialized internal controls, that is, known quantities of starting material that were co-amplified with the test template. These types of assays suffered, however, from primer incompatibilities and inconsistent results. Another approach was to add competitor templates at several known levels to assess the amount of test material by preferential amplification over a known amount of competitor.[25] These assays were also unreliable and inconsistent when test and internal control templates differed by more than 10-fold. They were most accurate with a 1:1 ratio of test and internal control, which required analysis of multiple dilutions of controls for optimal results.

A very useful modification of the PCR process is real-time or **quantitative PCR (qPCR).**[26,27] This method was first performed by adding ethidium bromide (EtBr) to a standard PCR. Because EtBr intercalates into double-stranded DNA and fluoresces, it tracks the accumulation of PCR products during the PCR in real time, that is, as it is made. Detectable fluorescence in earlier cycles of the amplification program indicates higher amounts of starting template, whereas fluorescence appearing in later cycles indicates lower amounts of starting template. This method more accurately reflects the amount of starting template. Furthermore, the quantitative measurements are performed with the ease and rapidity of standard PCR without the addition of competitor templates or multiple internal controls. The rationale for qPCR is illustrated in Figure 6.13. Graphing the PCR cycles (denaturation, annealing, extension) on the *x*-axis versus

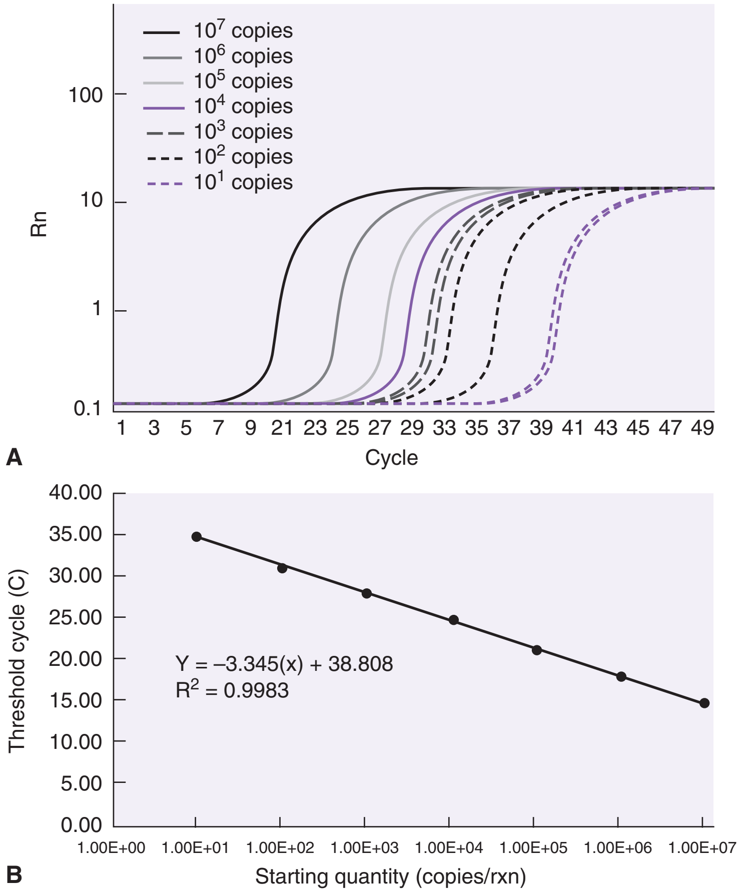


FIGURE 6.13 A plot of the accumulation of PCR product over 50 cycles of PCR. (A) A sigmoid curve. The generation of fluorescence occurs earlier with more starting template (solid lines) than with less (dotted lines). See Color Plate 1. The cycle number at which fluorescence increases over a set amount, or fluorescence threshold, is inversely proportional to the amount of starting material (B).

fluorescence generated (*y*-axis) yields an exponential curve where the number of copies = 2^N, with N being the number of PCR cycles. The curve looks similar to a bacterial growth curve, with a lag phase, an exponential (log) phase, a linear phase, and a stationary phase.

The exhaustion of reaction components and competition between PCR product and primers during the annealing step slow the PCR product accumulation after the exponential phase of growth until it finally plateaus. In contrast to qPCR, analysis of PCR product by the standard method occurs at the end of the PCR stationary phase **(endpoint analysis).** In the endpoint analysis, products of widely different amounts of starting template are tested at the plateau where they are all the same (observe the ends of the amplification curves shown in Fig. 6.13A). Using the fluorescent signal to detect the growing target copy number during the amplification process, analysis in qPCR is performed in the exponential phase of growth. Because the length of the lag phase is inversely proportional to the amount of starting template, fluorescence will reach exponential growth in early cycles when a lot of target is present; when less target is present, fluorescence will not reach exponential growth until later cycles. With serial fold dilutions of known positive standards, a relationship can be established between the starting target copy or cell number and the cycle number at which fluorescence crosses a threshold amount of fluorescence.

Advanced Concepts

The optimal threshold level is based on the background or baseline fluorescence and the peak fluorescence in the reaction. Instrument software is designed to set this level automatically. Alternatively, the threshold may be determined and set manually.

The PCR cycle at which sample fluorescence crosses the threshold is the **threshold cycle,** or **C_T**. Plotting the target copy number of the diluted standards against C_T for each standard generates the graph shown in Figure 6.13B. Once this relationship is established, the starting amount of an unknown specimen can be determined by the cycle number at which the unknown crosses the fluorescence threshold. This method is applied to the quantification of DNA targets (qPCR) and RNA targets in reverse-transcriptase qPCR (RT-qPCR). The template for RT-qPCR is cDNA.

The first approach to qPCR utilized EtBr, which is specific to double-stranded DNA. Dye is still used for routine qPCR, except that EtBr has been replaced by **SYBR green,** another dye specific to double-stranded DNA. The advantage of SYBR green is its specificity and robust fluorescence comparable to that of EtBr and its reduced toxicity. Both dyes bind and fluoresce specifically in the double-stranded DNA product of the PCR (Fig. 6.14).

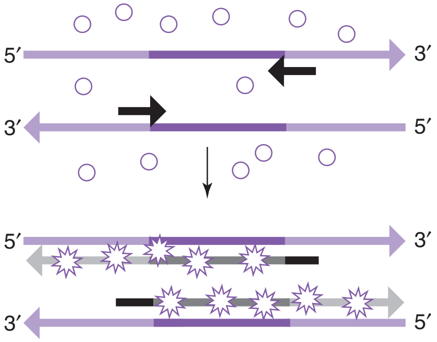


FIGURE 6.14 Non-sequence–specific dyes, such as EtBr and SYBR green, bind to double-stranded DNA products of the PCR. As more copies of the target sequence accumulate, the fluorescence increases.

The use of nonspecific dyes to measure the accumulation of product requires a clean PCR free of mispriming and primer dimers because these artifactual products will also generate fluorescence. More specific systems, examples of which are described later in the chapter, have been devised that utilize probes designed to generate fluorescence. The probes increase specificity by only yielding fluorescence when they hybridize to the target sequences.

Advanced Concepts

Fluorescence versus C_T is an inverse relationship. The more starting material, the fewer cycles are necessary to reach the fluorescence threshold. Samples that differ by a factor of 2 in the original concentration of target are expected to be 1 cycle apart, with the more dilute sample having a C_T 1 cycle higher than the more concentrated sample. Samples that differ by a factor of 10 (as in a 10-fold dilution series) would be 3.3 cycles apart. The slope of a standard curve made with 10-fold dilutions, therefore, should be –3.3.

TaqMan was developed from one of the first probe-based systems for quantifying cDNA by qPCR.[28] This method exploits the natural 5′ to 3′ exonuclease activity of *Taq* polymerase to generate a signal. An early version of the method reported by Holland et al. used radioactively labeled probe and measured activity by the release of radioactive cleavage fragments.[29] The TaqMan procedure measures the fluorescent signal generated by separation of fluorescent dye and **quencher,** a system developed by Lee et al. that used a probe composed of a single-stranded DNA oligonucleotide complementary to a specific sequence in the targeted region of the PCR template.[30] The probe is present in the reaction mix in addition to the specific primers that prime the DNA synthesis reaction. The probe is chemically modified at its 3′ end so that it cannot be extended by the polymerase. The single-stranded DNA TaqMan probe is covalently attached to a fluorescent dye on the 5′ end and another dye or nonfluorescent molecule that pulls fluorescent energy from the 5′ dye (quencher) on the 3′ end[31] (Fig. 6.15).

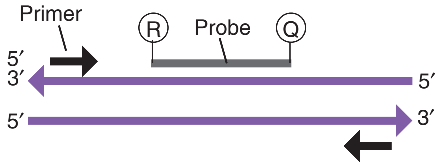


FIGURE 6.15 A TaqMan probe hybridizes to the target sequences between the PCR primer-binding sites. The probe is covalently attached to a fluorescent reporter dye (R) at the 5′ end and a quencher (Q) at the 3′ end.

Advanced Concepts

EtBR is a planar molecule that intercalates between the planar nucleotides in the DNA molecule. In doing so, it interferes with DNA metabolism and replication in vivo and is a mutagen. In contrast, SYBR green binds to the minor groove of the double helix without disturbing the nucleotide bases and thus does not upset DNA metabolism to the extent that EtBr does.

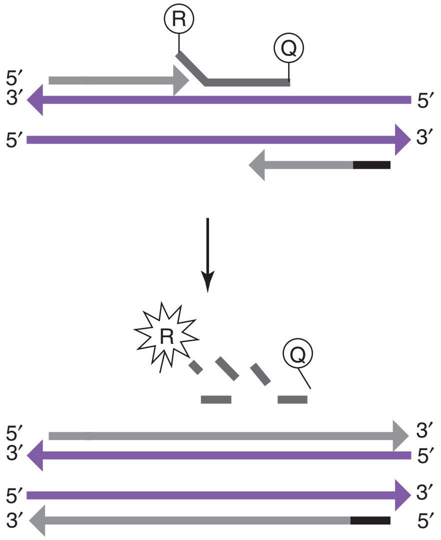


FIGURE 6.16 TaqMan signal fluorescence is generated when *Taq* polymerase extends the primers and digests the probe and releases the reporter from the vicinity of the quencher.

As the polymerase proceeds to synthesize DNA on the template to which the probe is hybridized, the natural exonuclease activity of *Taq* polymerase will degrade the probe into single and oligonucleotides, thereby removing the labeled nucleotide from the vicinity of the quencher and allowing it to fluoresce (Fig. 6.16). Excess probe is present so that with every doubling of the target sequences, more probe binds and is digested, and more fluorescence is generated. Probe design, like primer design, is important for a successful qPCR amplification, as are the characteristics of the polymerase enzyme.[32]

The 5′ end of a TaqMan probe is labeled with one of a number of dyes with different "colors," or peak emission wavelengths, of fluorescence, for example, FAM (6-carboxyfluorescein), TET (6-tetrachlorofluorescein), HEX (6-hexachlorofluorescein), JOE (4′, 5′-dichloro-2′, 7′-dimethoxy-fluorescein), Cy3, and Cy5 (indocarbocyanine). The probe is covalently bound at the 3′ end with a quencher, such as DABCYL (4-dimethylaminophenylazobenzoic acid) or TAMRA ([56]-carboxytetramethylrhodamine), or nonfluorescent quenchers, such as BHQ1, BHQ2 (Black Hole Quenchers), and Eclipse. In the TaqMan system, the quencher prevents fluorescence from the 5′ dye until they are separated during the synthesis reaction. As more copies of the template accumulate, more of the 5′ dye is accumulated throughout the reaction program.

Another probe-based detection system, **Molecular Beacons,** measures the accumulation of product at the annealing step in the PCR cycle.[33] The signal from Molecular Beacons is detectable only when the probes are bound to the template *before* displacement by the polymerase. Here the probe is chemically modified so that it is not degraded during the extension step. Molecular Beacons are designed with a target-specific binding sequence of approximately 25 bases flanked by a short, approximately 5-base-long inverted repeat that will form a stem and loop structure when the probe is not bound to the template. There is a reporter fluorophore (dye) at the 5′ end of the oligomer and a quencher at the 3′ end. Until the specific product is present, the probe will form a hairpin structure that brings the fluorophore

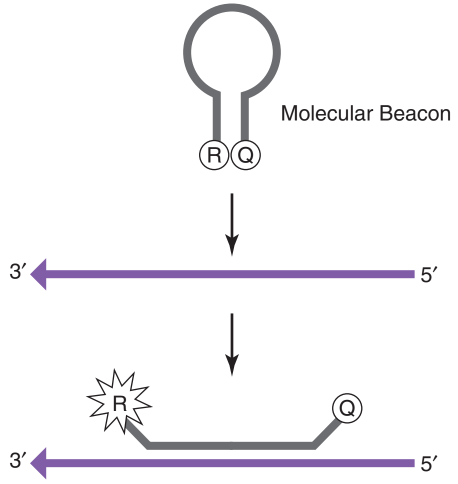


FIGURE 6.17 A Molecular Beacon probe contains target-specific sequences and a short, inverted repeat (~5 bp) that hybridizes into a hairpin structure. The 5′ end of the probe has a reporter dye (R), and the 3′ end has a quencher dye (Q).

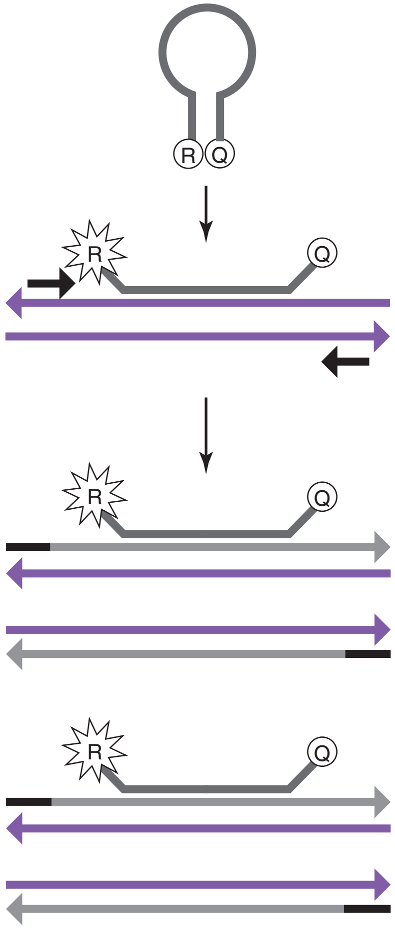


FIGURE 6.18 In the presence of target sequences, hybridization of the probe will open the hairpin, moving the quencher from the reporter and allowing signal fluorescence, which doubles with every doubling of target sequences.

in proximity with the quencher (Fig. 6.17). Fluorescence will occur on the binding of the probe to the denatured template during the annealing step (Fig. 6.18). When the primers are extended in the PCR, displacement of the probe by *Taq* will restore the hairpin (nonfluorescent) structure. Excess probe in the reaction mix will ensure binding to the increasing amount of target. The amount of fluorescence, therefore, will be directly proportional to the amount of template available for binding and inversely proportional to the C_T.

Scorpion-type primers are a variation of Molecular Beacons.[34,35] In contrast to free-labeled probes, the PCR product will be covalently bound to the dye. In this system, target-specific primers are tailed at the 5′ end with a sequence complementary to part of the internal primer sequence, a quencher, a stem-loop structure, and a 5′ fluorophore (Fig. 6.19). The fluorophore and the quencher are positioned so that they are juxtaposed when the hairpin in the primer is intact. After polymerization, the secondary structure of the primer is overcome by hybridization of the primer sequence with the target sequence, removing the fluorophore from the quencher. This intramolecular system generates a signal faster than the intermolecular Molecular Beacon strategy and may be preferred for methods requiring fast cycling conditions.[36] **Scorpions** also produce a PCR product that is covalently labeled with the fluorophore that can be further analyzed by capillary electrophoresis.

Another frequently used system, **fluorescent resonance energy transfer (FRET),** utilizes two specific probes, one with a 3′ fluorophore (acceptor) and the other with a 5′ catalyst for the fluorescence (donor), that bind to adjacent targets.[37] Examples of frequently used donor–acceptor pairs are fluorescein–rhodamine,

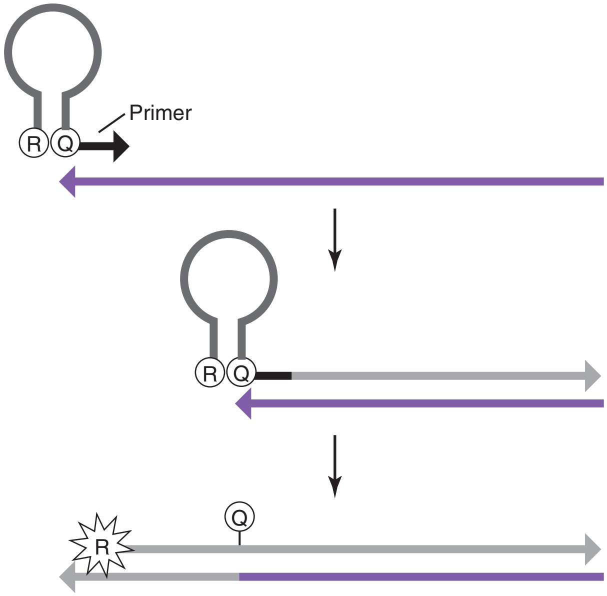


FIGURE 6.19 Scorpion primer/probes are primers tailed with Molecular Beacon-type sequences. After extension of the primer/probe, the target-specific sequences fold over to hybridize with the newly synthesized target sequences, separating the reporter (R) from the quencher (Q). An advantage of this system is the covalent attachment of fluorescent signal to the PCR product, which is useful for further analysis, such as size assessment by capillary electrophoresis.

fluorescein–(2 aminopurine), and fluorescein–Cy5. When the donor and acceptor are brought within 1 to 10 nm (1 to 5 bases) through specific DNA binding, excitation energy is transferred from the donor to the acceptor (Fig. 6.20). The acceptor then loses the energy in the form of heat or fluorescence emission. As with the Molecular Beacons, the more template available for binding of the probes, the more fluorescence will be generated. Quantitative PCR lends itself to several variations of technique, as exemplified previously. These techniques can be further modified, for example, using FRET probes with different sequences to distinguish types of organisms or to detect mutations. FRET probes are also part of methods that use melt curves to detect gene mutations. As with standard PCR, many such methods have been devised for a variety of applications.

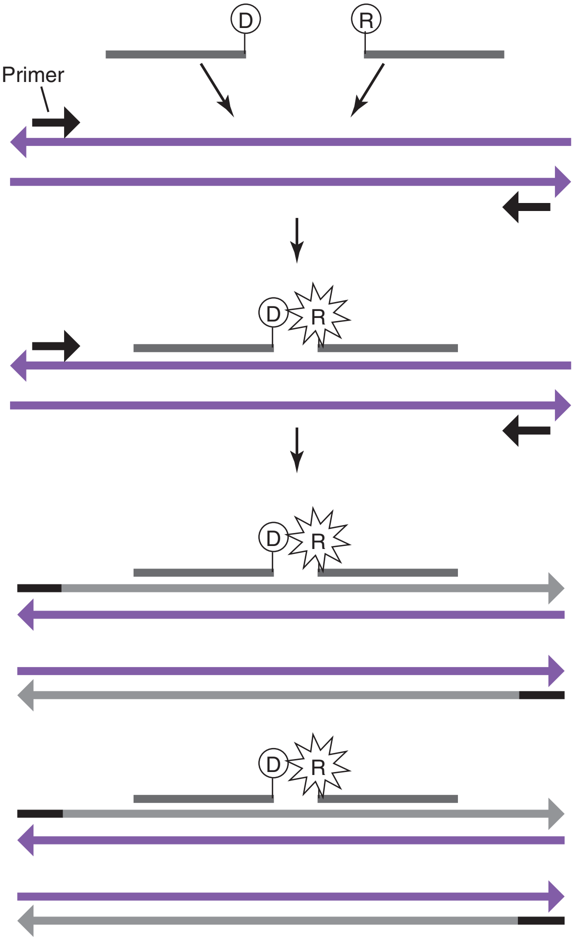


FIGURE 6.20 FRET probes are separate oligomers, one covalently attached to a donor fluor (D) and one to an acceptor or reporter fluor (R). The acceptor/reporter will fluoresce only when both probes are bound next to one another on the target sequences. As more target accumulates, more probes bind, and more fluorescence is emitted.

Advanced Concepts

Internal controls are used with qPCR assays to detect false-negative results in the event of amplification failure. Ideally, these controls should be RNA for RNA expression assays and plasmid for DNA copy-number analyses.

Advanced Concepts

The ease of use and flexibility of qPCR has led to the implementation of a variety of instruments, reagents, and analysis methods that have sometimes produced conflicting results. The Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines recommend the reporting of performance characteristics for any qPCR assay used for publication.[38,39] Performance metrics include target specificity, PCR efficiency, limit of detection, precision, dynamic range, and either primer sequences or amplicon context. These measurements can also be used to evaluate reagents. For clinical laboratories, they are included in the validation of qPCR methods to be used for medical laboratory testing.

Transcription-Based Amplification Systems

In a transcription-based amplification system (TAS), RNA is the usual target instead of DNA. A DNA copy is synthesized from the target RNA, and then transcription of the DNA produces millions of copies of RNA. There are a number of commercial name variations of this **process: transcription-mediated amplification (TMA), nucleic acid sequence–based amplification (NASBA),** and **self-sustaining sequence replication (3SR).**

Kwoh and colleagues developed the first TAS in 1989.[40] TAS differs from other nucleic acid amplification procedures in that RNA is the target as well as the primary product. In the original method of TAS, a primer carrying the binding site for RNA polymerase at its 5′ end and complementary to sequences in the target RNA is added to a sample of target RNA. The primer anneals, and reverse transcriptase makes a DNA copy of the target RNA. Heat is used to denature the DNA–RNA hybrid, and a second primer binds to the cDNA and is extended by reverse transcriptase, producing double-stranded DNA. RNA polymerase derived from the bacteriophage T7 then transcribes the cDNA, producing hundreds to thousands of copies of RNA. The transcribed RNA can then serve as target RNA to which the primers bind and synthesize more cDNA.

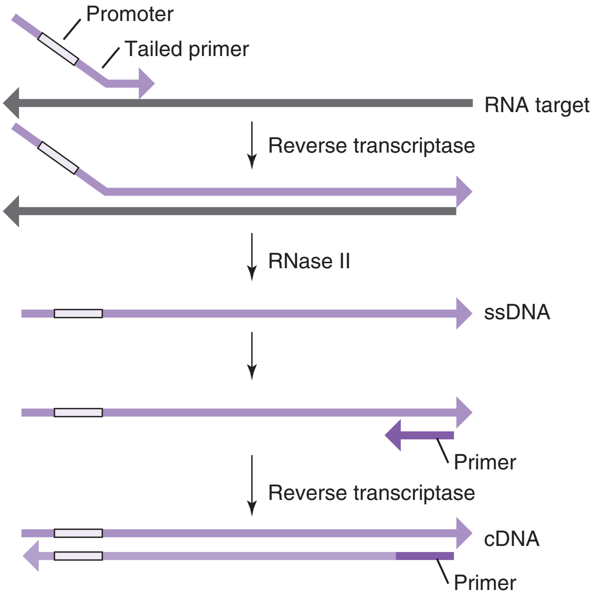


FIGURE 6.21 The first step in transcription-based amplification is the production of a complementary double-stranded DNA copy of the RNA target. Synthesis is performed by reverse transcriptase, which extends a primer that is tailed with an RNA polymerase-binding site (promoter) sequence (purple). The RNA:DNA hybrid is digested with RNase H, leaving the single-stranded DNA, which is converted to a double strand with a complementary primer. The DNA product will have a promoter sequence at one end.

The original TAS procedure as just described had the disadvantage that a heating step was required to denature the intermediate RNA:DNA hybrid product. The heat also denatured the enzymes so that fresh enzyme had to be added after each denaturation step. The process was simplified with the addition of RNase H derived from *E. coli* (Fig. 6.21). RNase H degrades the RNA from the intermediate hybrid, eliminating the heating step. Thus, after synthesis of the DNA copy by reverse transcriptase, the RNA strand is degraded by RNase H. Binding of the second primer and extension of the primer, producing double-stranded DNA by reverse transcriptase, is followed by transcription of the cDNA with T7 RNA polymerase (Fig. 6.22).

An additional modification and simplification of the procedure came about with the discovery that the reverse transcriptase derived from avian myeloblastosis

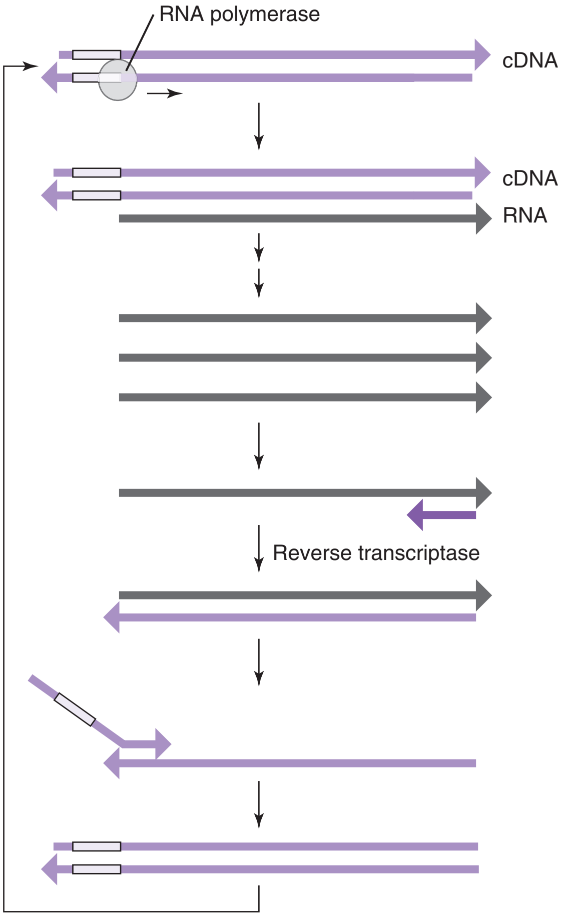


FIGURE 6.22 The cDNA produced in the first step (Fig. 6.21) serves as a template for RNA polymerase. Many copies of RNA are synthesized, which are primed by a complementary primer for synthesis of another RNA:DNA hybrid. After RNase H degrades the RNA strand, the primer tailed with the promoter sequences synthesizes another template, cycling back into the system as a template for RNA polymerase.

virus (AMV) has inherent RNase H activity. Thus, TAS can be run with only two enzymes, AMV reverse transcriptase and T7 RNA polymerase.

TAS has some advantages over PCR and other amplification procedures. First, in contrast to PCR, TAS is an isothermal process, negating the requirement for thermal cycling and heat-stable enzymes to drive the reactions. Second, targeting RNA allows for the direct detection of RNA viruses, for example, hepatitis C virus and human immunodeficiency virus (HIV). Even targeting the RNA of organisms with DNA genomes, such as *Mycobacterium tuberculosis,* is more sensitive than targeting the DNA because each bacterium, for example, has multiple copies of RNA, whereas it has only one copy of DNA.

TMA, like NASBA, can also start with a DNA target.[41] For DNA, the sample is heated to denature the DNA, and the first primer anneals and is extended by reverse transcriptase (which also has DNA-dependent DNA polymerase activity in addition to having RNA-dependent DNA polymerase activity). The RNA strand is removed, and the second primer binds and is extended. The DNA product has also incorporated the T7 RNA polymerase binding site, which is on the 5′ end of the first primer. Thus, T7 RNA polymerase transcribes the newly replicated DNA into hundreds to thousands of RNA copies. Detection of *M. tuberculosis* in smear-positive respiratory samples, *Chlamydia trachomatis* in genital specimens, and HIV and cytomegalovirus (CMV) quantitation in blood are a few early applications of TAS.

Genomic Amplification Methods

In addition to amplification of single genes or transcripts, methods have been devised to amplify all regions of input DNA, or whole genomes. Whole-genome amplification (WGA) methods are designed to survey all genes or transcripts of an organism for the purpose of typing of microorganisms or screening for particular genetic lesions from limiting samples. Genomic amplification differs from standard amplification methods in that the primers used are not complementary to specific genetic regions; rather, replication initiates at random sites throughout the input nucleic acid (Fig. 6.23).[42]

Genomic amplification provides the ability to generate a lot of information from a very small amount of

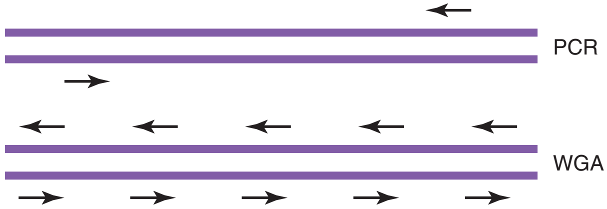


FIGURE 6.23 In contrast to PCR, the object of whole-genome amplification (WGA) is to copy all regions of a genomic template. The WGA primers may be of random sequence or directed to highly repeated sequences in the genome.

starting material. In biological samples especially, the target nucleic acid is not always clean, accessible, or in optimal amounts for complex analyses, such as sequencing or microarray methods. The following methods are approaches to genomic amplification applied to the typing of microorganisms and generation of multiple genomic copies for genetic studies.

Whole-Genome Amplification

WGA can be performed using degenerative primers that prime random synthesis throughout the target genome or by nick translation. Degenerative primers are synthetic single-stranded DNA sequences that vary with some frequency at each position where N = A, T, G, or C. In primer extension preamplification (PEP), this variation occurs throughout the primer: $N_{15–16}$. In degenerative oligonucleotide primer (DOP) amplification, the variable sequences are flanked by a highly repeated sequence in the source DNA, for example, for human DNA: $CCGACTCGAGN_6ATGTGG$. In tagged-PCR (T-PCR), primers have a 5′ sequence complementary to repetitive DNA followed by a random sequence of variable length: $CTCACTCTCAN_X$.

Alternatively, whole-genome amplification can be achieved by introducing single-strand breaks or nicks in the genomic DNA, with amplification primed by short random sequences (hexamers) or the 3′ ends of the breaks.[43] This is multiple-displacement amplification because there is removal or denaturation of double-stranded DNA as extension proceeds. Amplification is performed using Phi29 DNA polymerase, a highly processive enzyme—one that can displace and copy thousands of bases without losing contact with the template.

WGA has multiple applications, including **comparative genomic arrays,** detection of single-nucleotide polymorphisms, and analysis of minimal starting material, such as DNA in plasma, single-cell analysis, and ancient DNA samples.

Emulsion PCR

Emulsion PCR is designed to simultaneously amplify thousands of specific templates in a single reaction, producing a set of specific products, or library.[44] In this method, the ends of fragmented template DNA are ligated to universal primer sequences and, along with the other components of the PCR reaction (buffer, polymerase enzyme, forward and reverse primers, dNTPs), are added to an oil surfactant mixture (Span 80, Tween 80, Triton X-100, and mineral oil; Fig. 6.24). With stirring, an emulsion is formed of 1 to 10 billion droplets, each of which can contain a single template in a tiny volume of reaction buffer. When the emulsion is subjected to an amplification program, the droplets will act as single reaction chambers producing many copies of a specific sequence. After amplification, the emulsion is broken. The pooled products are used for a variety of genomic study applications, such as sequencing, array studies, haplotyping, or detection of rare mutations.

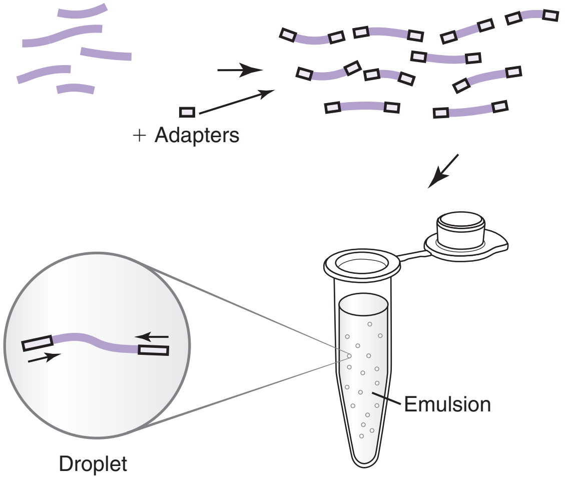


FIGURE 6.24 Whole-genome DNA is fragmented, and the fragments are ligated to adapters that are complementary to one set of forward and reverse PCR primers. The adapted templates, primers, and other PCR reagents in aqueous solution form an emulsion with oil such that each aqueous droplet in the emulsion contains a single template. Each of thousands of oil droplets then serve as reaction chambers for unique products.

Solid-phase emulsion PCR was developed for next-generation sequencing and other high-throughput technologies that have driven the development of equally high-throughput PCR technologies. In this method, templates ligated to a universal primer sequence are mixed with bead-immobilized primers and other reaction components in the water-in-oil emulsion (Fig. 6.25). Aqueous droplets in the oil ideally would capture a single template and a single primer-coupled bead. During the PCR reaction, the primers are extended, producing a

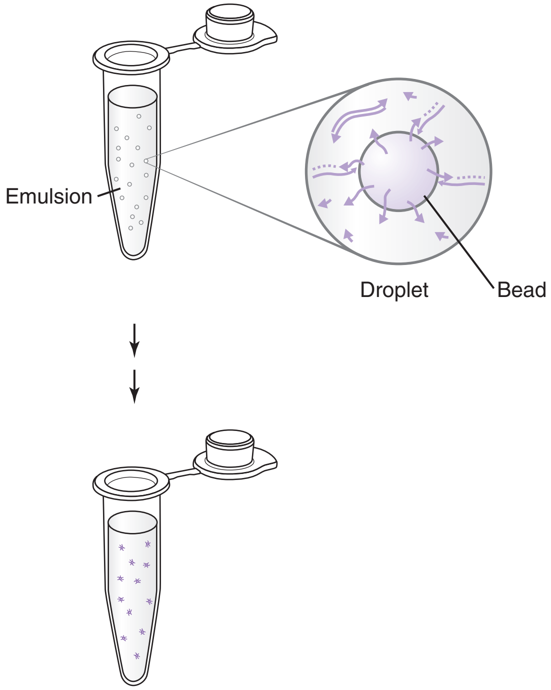


FIGURE 6.25 For solid-phase ePCR, the template is prepared as described in Figure 6.25. Many copies of one primer are covalently attached to a microsphere (bead). In the emulsion droplets, a single template will be copied into immobilized products on the beads. After the PCR reaction, the emulsion is broken, the beads are released, and the complementary strands are washed away, leaving many copies of immobilized single strands on each bead.

double-stranded copy from the template attached to the bead. When the emulsion is broken, the PCR products are denatured, washing away the complementary strand and leaving the single-stranded sequencing templates attached to the beads. These templates can then be subjected to massive parallel (next-generation) sequencing technologies.

Surface Amplification (Bridge PCR)

Surface amplification is an isothermal, genomic PCR method used in high-throughput sequencing technologies. In this method, forward and reverse primers are immobilized on a solid support in a flow cell (Fig. 6.26). Fragmented or preamplified template is denatured, and those single strands complementary to the immobilized primers will anneal under nondenaturing buffer conditions. After extension of the primer, denaturing conditions are introduced by the addition of formamide to the reaction mixture. The template and reaction components are then removed by washing, and annealing conditions are reestablished chemically, allowing the attached copy of the template to anneal to the immobilized reverse primer. The reverse primer is extended, forming a bridge between the two immobilized primers. The process is repeated for 35 cycles, producing approximately 1,000 copies of the template sequence in a tiny region of the surface. The double-stranded bridges are then denatured, resulting in a localized clone or **polony** of single-stranded complementary molecules. To avoid re-annealing, a cleavable site located on one primer (either forward or reverse) allows removal of one of the complements (Fig. 6.27). For DNA sequence applications, the remaining strand is chemically blocked at the free 3′ end using terminal DNA transferase and a dideoxynucleotide molecule that cannot be extended. Sequencing primer is then annealed to the template for sequencing.

Arbitrarily Primed PCR

In arbitrarily primed PCR, also known as **randomly amplified polymorphic DNA** or random amplification of polymorphic DNA (RAPD), short primers (10 to 15 bases) with random sequences are used to amplify arbitrary regions in genomic DNA under low-stringency conditions.[45,46] With this method, PCR products are generated without knowing the sequence of the target or targeting a specific gene. In contrast to standard PCR where only one or a few known products are generated, multiple products are generated depending on how many times a short sequence appears in the genome (Fig. 6.27). Arbitrarily primed PCR has been used primarily in the epidemiological typing of microorganisms. Similar band patterns obtained from performing PCR with the same arbitrary primers indicate that two organisms are the same or similar. A disadvantage of this method, however, is that reproducibility between runs is not very good, such that two organisms that had the same PCR product pattern on one day could have two different patterns and look like two different organisms when amplified on another day. Other methods such as mass spectrometry and rRNA sequencing provide more

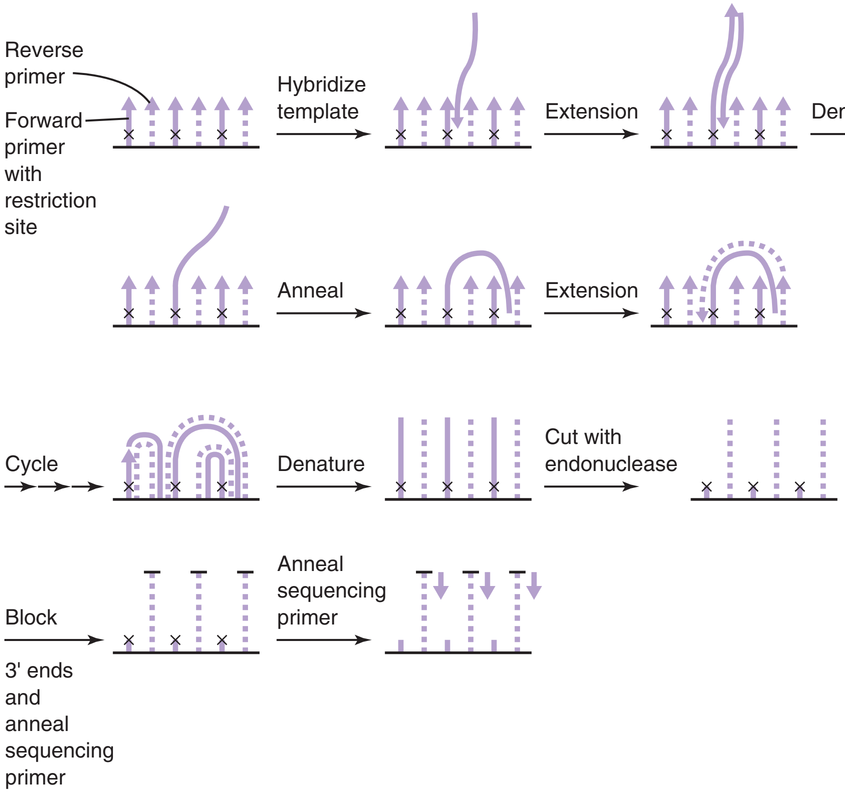

Reverse primer
Forward primer with restriction site
Hybridize template
Extension
Denature
Anneal
Extension
Cycle
Denature
Cut with endonuclease
Block 3' ends and anneal sequencing primer
Anneal sequencing primer

FIGURE 6.26 Bridge PCR produces immobilized single-stranded products from immobilized forward and reverse primers. One primer is designed to contain a single-strand endonuclease recognition site. After amplification, the PCR products form bridges from the forward to the reverse primers so that upon denaturation, the single strands are also immobilized. Endonuclease digestion will remove one complement of single strands so that the remaining population of immobilized single strands is identical. The strands are then subjected to further analysis, such as sequencing.

reproducible typing results; however, RAPD offers a relatively inexpensive alternative.

PROBE AMPLIFICATION

In **probe amplification** procedures, the number of target nucleic acid sequences in a sample is not changed. Rather, synthetic probes that are specific to the target sequences bind to the target where the probes themselves are amplified. There are three major procedures that involve the amplification of probe sequences: ligase chain reaction (LCR), strand displacement amplification (SDA), and Qβ replicase.

Ligase Chain Reaction

LCR was a method for amplifying synthetic primers/probes complementary to target nucleic acid. Similar to PCR, the entire target sequence had to be known in order to prepare the oligonucleotide primers for LCR. In PCR, there is a distance between the primers of hundreds to thousands of bases that is part of the amplified sequence. In LCR, by contrast, the primers are bound immediately adjacent to each other. Instead of DNA polymerase synthesizing complementary DNA by extending the primers as occurs in PCR, DNA ligase was used in LCR to ligate the adjacent primers together. The ligated primers then served as a template for the annealing and ligation of

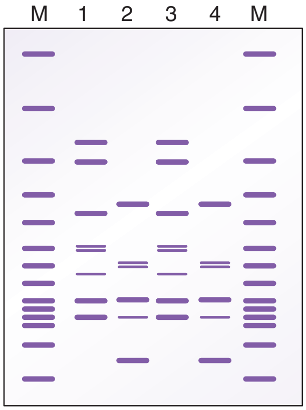


FIGURE 6.27 RAPD or arbitrarily primed PCR yields many products per PCR reaction. The band pattern produced by these fragments reflects the nature of the DNA template. Lanes 1 and 3 are products from very similar templates, which are different from those in lanes 2 and 4. *M,* molecular-weight markers.

additional primers. Because the product of LCR was ligated primer, LCR was a method of probe amplification rather than target amplification because the copy number of target molecules did not change.

LCR required a thermal cycler to change the temperature to drive the different reactions. In LCR the reaction was heated to denature the template. When the temperature was cooled, the primers annealed if the complementary sequence was present, and a thermostable ligase joined the two primers (Fig. 6.28). Even a 1-bp mismatch at the ligation point prevented ligation of the primers. LCR was used to detect point mutations in a target sequence. The DNA mutation that occurs in the beta globulin of patients with sickle cell disease, as compared with normal beta globulin, was one of the first applications of LCR.[47] LCR in the clinical laboratory has mostly been replaced by other methods; however, new applications of the technology have been proposed using FRET for detection of DNA single-nucleotide changes,[48] quantification of RNA,[49] and DNA methylation analysis.[50]

Strand Displacement Amplification

Strand displacement amplification (SDA) is an isothermal amplification process; that is, after an initial denaturation step, the reaction proceeds at one temperature.[51]

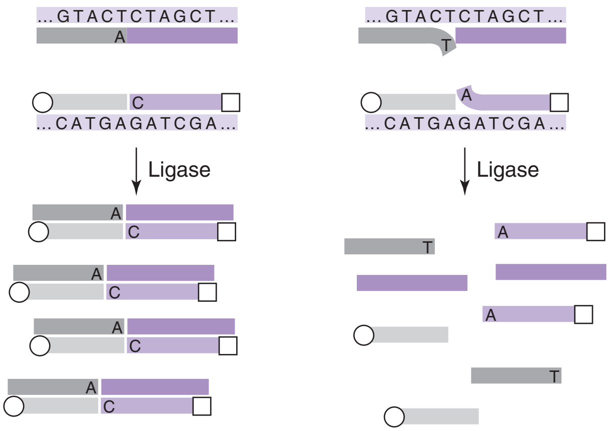


FIGURE 6.28 Ligase chain reaction generates a signal by repeated ligation of probes complementary to specific sequences in the test DNA. One complementary oligomer is covalently attached to biotin for immobilization *(square),* and one has a signal-producing molecule *(circle).* The two oligomers will be ligated together only if the sequence of the target is complementary *(left).* The oligomers captured on a solid substrate by streptavidin will generate a signal. If the sequence of the target is not complementary *(right),* the captured probe will not yield a signal.

In SDA the major amplification products are the probes. There are two stages to the SDA process. In the first stage (target generation), the target DNA is denatured by heating to 95°C. At each end of the target sequence, two primers bind close to each other, an outer and an inner primer (Fig. 6.29). The inner primers have a 5′ tail containing a recognition sequence for the restriction endonuclease enzyme, *Hinc*II. Exonuclease-deficient DNA polymerase derived from *E. coli* DNA polymerase I extends all of the primers, incorporating a modified nucleotide, 2′-deoxyadenosine 5′-O-(1-thiotriphosphate) (dATPaS), into the nucleotide mix. As the outer primers are extended, they displace the products formed by the extension of the inner primers. A second set of outer and inner primers then binds to the displaced inner primer products, and DNA polymerase extends the complementary primers, producing four double-stranded products (probes). These probes are the target DNA for the next stage of the process.

The second stage of the reaction is the exponential probe amplification phase using *Hinc*II (Fig. 6.30). When the restriction enzyme is added to the double-stranded

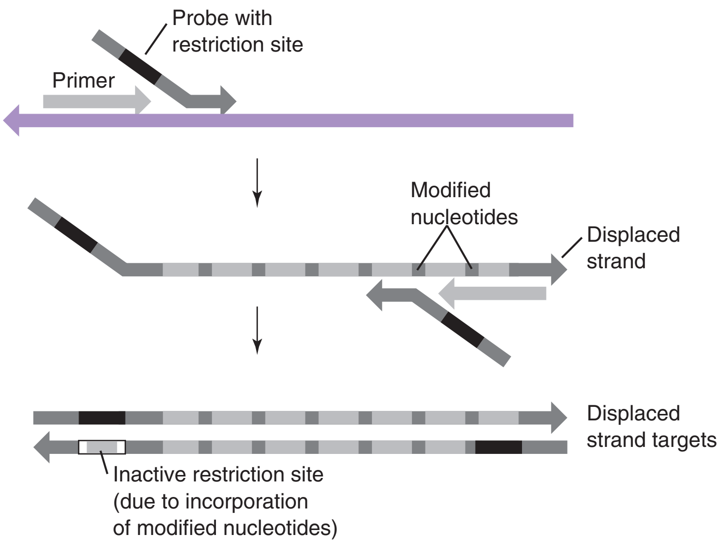

Probe with restriction site
Primer
Modified nucleotides
Displaced strand
Displaced strand targets
Inactive restriction site (due to incorporation of modified nucleotides)

FIGURE 6.29 The first stage of SDA is the denaturation of the double-stranded target and annealing of primers and probes tailed with sequences including a restriction enzyme site (only one strand of the initial target is shown). A second reaction copies the probe, incorporating dATPαS and thereby inactivating the restriction site on the copied strand. This species is the target for amplification in the second stage of the reaction.

probe DNA, only one strand of the probe will be cut due to the dATPαS introduced in the extension reaction. This forms a nick in the DNA that is extended by DNA polymerase, simultaneously displacing the opposite strands. The nicking and extension form single-stranded probes and regenerate the restriction site. The nicking/extension reaction can repeat on the initial probe as well as the probes generated in the reaction. With the enzyme nicking reaction, strand displacement does not require denaturation of the double-stranded probes. Thus, the iterative process takes place at about 52°C without temperature cycling. Addition of a fluorogenic probe to the reaction produces a fluorescent signal that corresponds to the amount of amplified target.

The SDA process was first widely applied to detection of *M. tuberculosis*.[52] Methods have been designed to test for *M. tuberculosis, C. trachomatis,* and *Neisseria gonorrhoeae*.

Advanced Concepts

For SDA of RNA or other single-stranded targets, the initial heat denaturation step is not necessary. The inner primer tailed with the restriction site is annealed to the 3′ end of the target, forming a double-stranded probe that will enter the iterative restriction/strand displacement reactions.

Qβ Replicase

Qβ replicase is another method for amplifying probes that have specificity for a target sequence. The method is named for the major enzyme that is used to amplify probe sequences. Qβ replicase is an RNA-dependent RNA polymerase from the bacteriophage Qβ.[53] The target nucleic acid in this assay can be either DNA (which must first be denatured) or RNA.

The target nucleic acid is added to a reaction mix containing reporter probes, which are RNA molecules (midivariant RNA) that have specificity for the target sequence and also contain a promoter sequence that is recognized by the Qβ replicase. The reporter probes are allowed to hybridize to the template. The template with bound probes is immobilized using polyG-tailed capture probes (Fig. 6.31, capture probe A) and magnetic beads covalently attached to polyC sequences. With a magnet applied to the outside of the wells, the unbound reporter molecules are washed away. The template–probe complexes are released from the polyC magnetic bead by denaturation and hybridized to a polyA capture probe (capture probe B). The complex is then hybridized to a polyT paramagnetic bead. After a series of washes to remove unbound reporter probe, the template–probe complex is again released from the magnetic bead.

For the amplification step, the midivariant RNA probe-bound template is mixed with the Qβ replicase, which replicates the midivariant RNA molecules. This

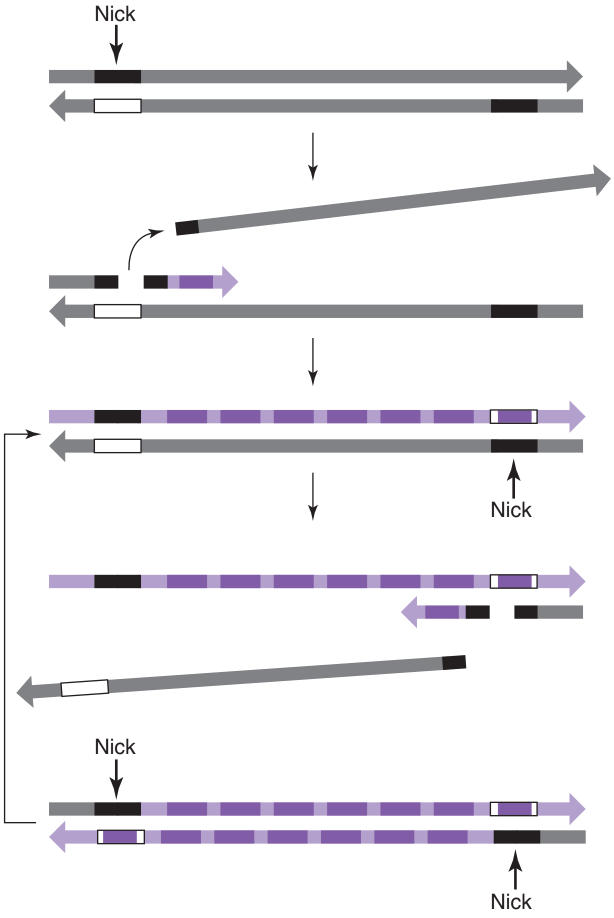

Nick
Nick
Nick
Nick

FIGURE 6.30 In the second phase of SDA, the target sequence is nicked by the restriction enzyme, generating a substrate for the polymerase, which extends the nick, displacing the opposite strand, and regenerates a new template for nicking. Copies of the displaced strands collect as the reaction proceeds. The reaction cycles by the strands are nicking and extension, without requirement for heat denaturation of the double-stranded DNA template.

replication is very efficient; it generates 10^6 to 10^9 RNA molecules (probes) in less than 15 minutes. Because so many RNA molecules are produced, product detection can be achieved by colorimetric as well as real-time fluorogenic methods. Qβ replicase has been replaced by other methods for most medical lab applications. It was used primarily to amplify the nucleic acid associated with infectious organisms, particularly mycobacteria, *Chlamydia,* HIV, and CMV.

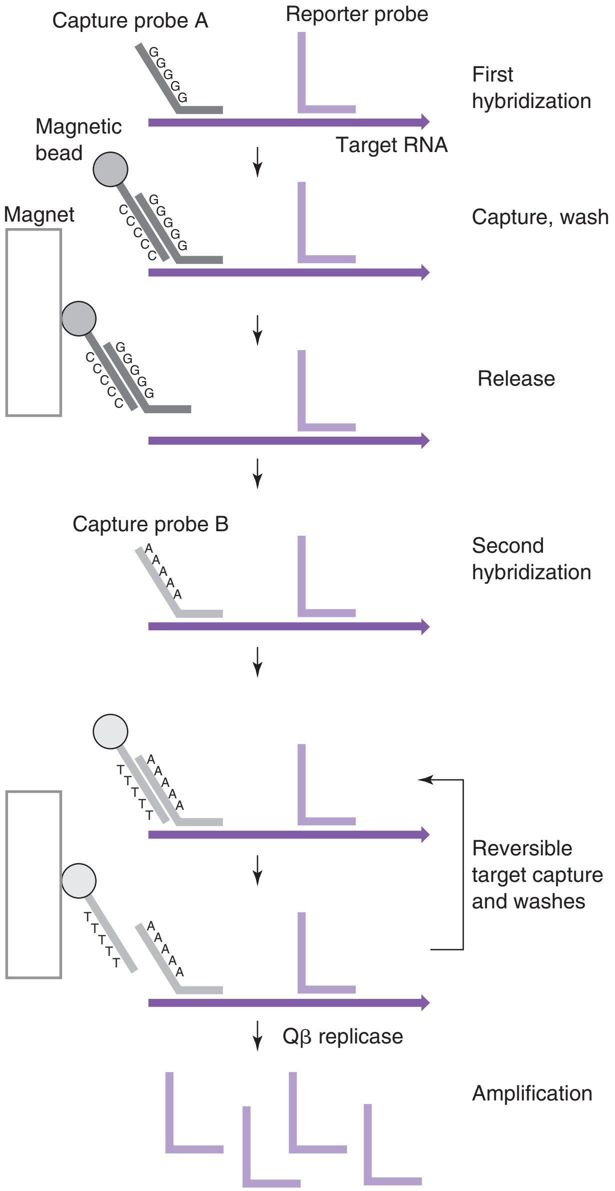

Capture probe A
Reporter probe
G G G G G
First hybridization
Magnetic bead
Target RNA
Magnet
C C C C C
Capture, wash
Release
Capture probe B
A A A A
Second hybridization
T T T T T
Reversible target capture and washes
Qβ replicase
Amplification

FIGURE 6.31 The Qβ replicase method proceeds through a series of binding and washing steps. Probe bound to the purified template is then amplified by Qβ replicase. The resulting RNA can be detected by fluorometry using propidium iodide as a fluorescent label of the synthesized probe or by chromogenic methods.

SIGNAL AMPLIFICATION

In **signal amplification** procedures, there is no change in the number of target or probe sequences; instead, large amounts of signal are bound to the target sequences that are present in the sample. Signal amplification procedures are inherently better at quantifying the amount of target sequences present in the clinical sample. Several signal amplification methods are available commercially.

Branched DNA Amplification

In **branched DNA (bDNA)** amplification,[54] a series of short oligomer probes is used to capture a single target nucleic acid molecule. Additional extender probes bind to the target nucleic acid and then to multiple reporter molecules, loading the target nucleic acid with signal.

For the bDNA signal amplification procedure, target nucleic acid released from the cells is denatured (if DNA is the target; this method also works with RNA). The target nucleic acid binds to capture probes that are fixed to the plate well (Fig. 6.32). Extender or preamplifier probes then bind to the captured target. The extender probes have sequences that are complementary to sequences in the target molecules as well as to sequences in amplifier probes.

In the first-generation assay, the extender probes bind to a tree-like bDNA amplifier probe, which in turn binds multiple alkaline phosphatase-labeled nucleotides. Eight multimers or amplifiers, each with 15 branches, bind to each extender probe bound to the target. In the second- and third-generation assays, the extender probes bind preamplifiers, which in turn bind 14 to 15 amplifiers, each with the capacity to bind multiple alkaline phosphatase–labeled oligonucleotides (Fig. 6.33). Dioxetane is added as the substrate for the alkaline phosphatase, and chemiluminescence is measured in a luminometer. This system has a detection limit of about 50 target mol/mL.[55]

There are several advantages to the bDNA method. First, in the bDNA assay there is less risk of carryover contamination resulting in a positive test than in PCR.[56] Second, multiple capture and extender probes can be incorporated that detect slightly different target

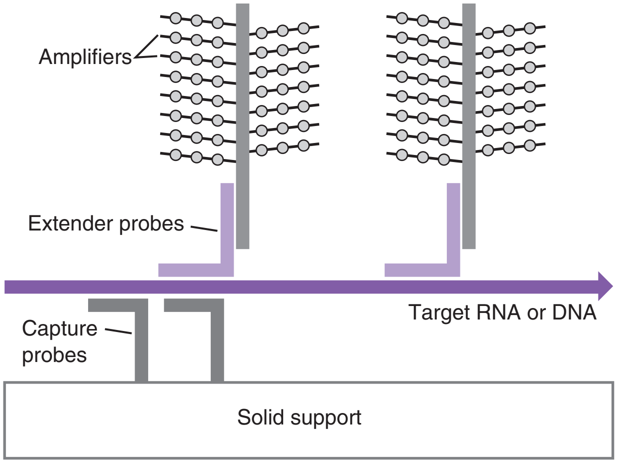


FIGURE 6.32 Branched DNA signal amplification of a single target. The target is captured or immobilized to a solid support by capture probes, after which extender probes and blocking probes create a stable cruciform structure with the amplifiers. Each amplifier has hybridization sites for 8 to 14 branches, which in turn bind substrate molecules for alkaline phosphatase.

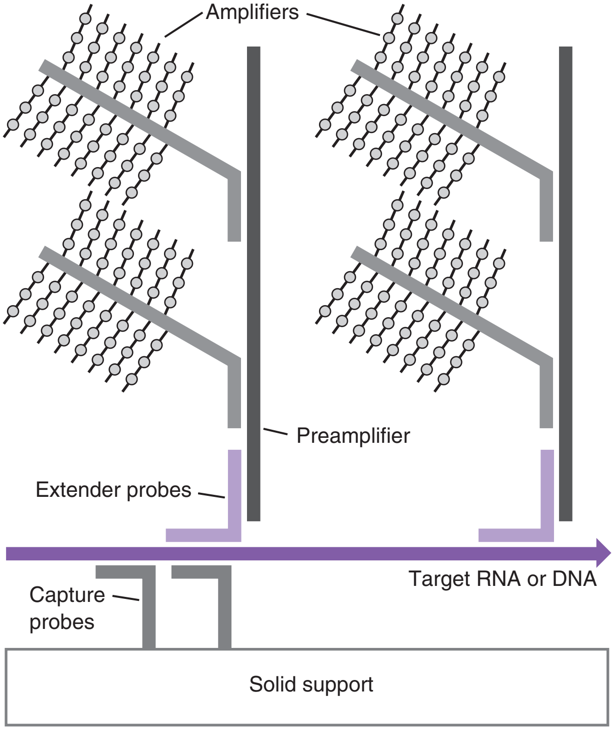


FIGURE 6.33 Second-generation bDNA assays use extender probes that bind multiple amplifiers, increasing the signal intensity and improving limits of detection.

sequences, as occurs with different isolates of hepatitis C virus and HIV. By incorporating different probes that recognize slightly different sequences, multiple genotypes of the same virus can be detected by the same basic system. Finally, the requirement for probes to bind multiple sequences in the same target increases the specificity of the system. It is highly unlikely that all of the required probes would bind nonspecifically to an unrelated target and produce a signal. The bDNA signal amplification assay has been applied to the qualitative and quantitative detection of hepatitis B virus, hepatitis C virus, and HIV-1. By replacing the plate support with beads, the assay has been combined with the bead array technology to provide a multiplex system that can detect 100 different targets in a single sample.[57]

Hybrid Capture Assays

The **hybrid capture** assays were marketed primarily for the detection and molecular characterization of human papillomavirus (HPV) in genitourinary specimens.[58,59] They are also used for the detection of hepatitis B virus and CMV. For these assays, target DNA released from cells is bound to single-stranded RNA probes (Fig. 6.34). The DNA:RNA hybrid has a unique structure that is recognized by antibodies. Antibodies bound to the surface of a microtiter well then capture the DNA:RNA hybrids. Double-stranded DNA or single-stranded RNA will not bind to these antibodies. Captured hybrids are detected by the binding of alkaline phosphatase–conjugated secondary antibodies to the DNA:RNA hybrid antibodies in a typical sandwich assay. A light-producing substrate for the alkaline phosphatase is added, and chemiluminescence is measured. The hybrid capture assay is considered a signal amplification assay because the amount of target DNA is not amplified; rather, the DNA is isolated bound to RNA and is recognized by multiple antibodies to the target/probe hybrid molecule.

Cleavage-Based Amplification

Cleavage-based amplification detects target nucleic acids by using a series of overlapping probes that bind to the target DNA. Cleavase is a bacterial enzyme that recognizes overlapping sequences of DNA and makes a cut (cleaves) in the overlapping region. In vivo, this activity is most likely important in repairing DNA. Applications for this form of amplification include DNA polymorphisms, such as factor V Leiden mutation detection,[60,61] and infectious diseases, such as hepatitis C virus (HCV) and HPV genotyping.[62,63]

To start the amplification, the target nucleic acid is mixed with invader and signal probes (Fig. 6.35). The invader probe and the signal probes bind at the target, with the 5′ end of the signal probe overlapping with the invader probe. Cleavase recognizes this triple-helix overlap and cleaves the signal probe, which can act as an invader probe in the next step of the reaction. In the second step, a FRET probe is added that has sequences complementary to the cleaved signal probe. The 5′ end of the FRET probe has a reporter molecule that is located in proximity to a quencher molecule. As a result, the intact FRET probe does not produce signal. The signal probe (now an invader probe) binds to the FRET probe, producing an overlapping region that is recognized by Cleavase. When Cleavase cuts the FRET

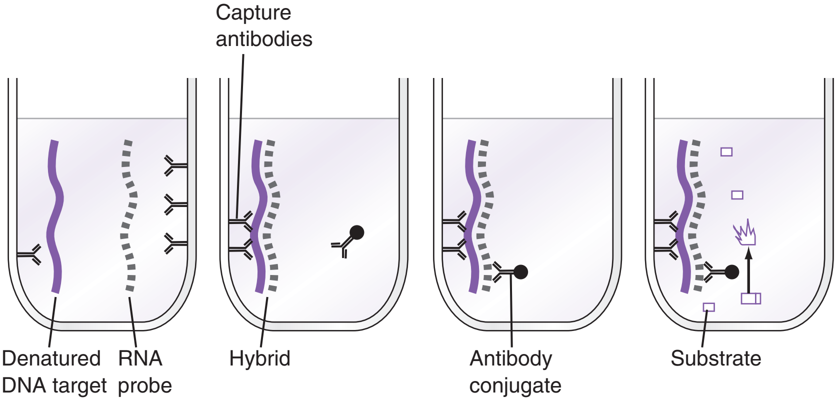


FIGURE 6.34 Hybrid capture starts with hybridization of the RNA probe to the denatured DNA target. The RNA:DNA hybrid is then bound by hybrid-specific immobilized antibodies. A secondary antibody bound to alkaline phosphatase generates signal in the presence of a chemiluminescent substrate *(right)*.

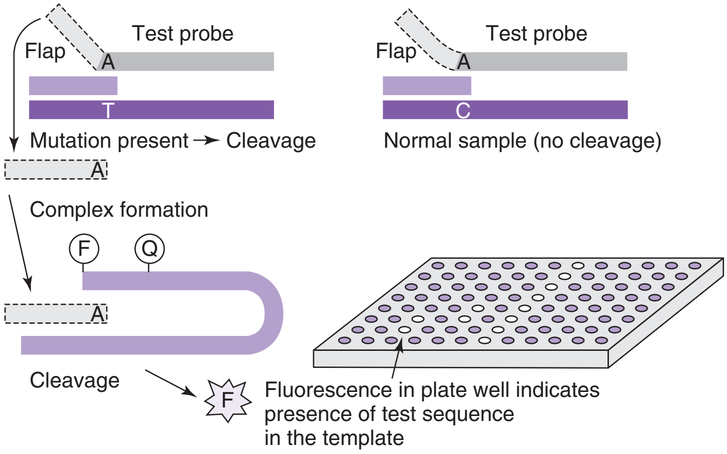

Flap
Test probe
A
T
Mutation present → Cleavage
A
Flap
Test probe
A
C
Normal sample (no cleavage)
Complex formation
F
Q
A
Cleavage
F
Fluorescence in plate well indicates presence of test sequence in the template

FIGURE 6.35 Invader assays exploit the substrate requirements of the enzyme. A cleavable substrate is formed by the hybridization of two probes and template *(top left)*. This structure will not form if the probe does not match the template *(top right)*. If the proper substrate is formed, the enzyme removes part of one probe, and that fragment forms another cleavable complex with the fluorescently labeled reporter probe *(bottom left)*. Repeated binding and cleavage amplify the signal. In a plate format, only those wells with a template complementary to the test probe will produce fluorescent signal.

probe in the overlapping region, it releases the reporter molecule from the quencher, resulting in the production of signal. The amount of signal can be quantified and related directly to the amount of target molecules in the sample. These reactions are carried out in a 96-well plate format, and the signal is detected on a plate reader.

Cycling Probe

In the **cycling probe** method of amplification, target sequences are detected using a synthetic probe consisting of sequences of DNA and RNA arranged in a DNA–RNA–DNA sandwich sequence carrying a **reporter dye** at one end and a quencher dye at the other. After the probe binds to the target nucleic acid (Fig. 6.36), RNase H cleaves the RNA from the middle of the probe. The loss of the RNA sequences lowers the hybridization temperature of the probe, and the DNA portions of the probes leave the template. When the probe is released, the reporter and quencher dye are separated, allowing fluorescence to escape from the reporter. The template remains available for additional probe hybridization. Because the cyclic denaturation of the probe depends on the digestion of the RNA portion, the method is isothermal. The amount of fluorescence from the reporter dye (produced when the target is present) is measured as an indication of the presence of target molecules. Alternatively, the presence of chimeric probes that remain when target sequences are not present can also be measured. This method has been used to detect genes associated with antimicrobial resistance in bacteria, such as methicillin resistance **(*mecA*)** in *Staphylococcus aureus* and vancomycin resistance (*vanA* and *vanB*) in *Enterococcus,*[64-66] and in the detection of other organisms, such as herpesvirus and *Histoplasma capsulatum*.[67,68]

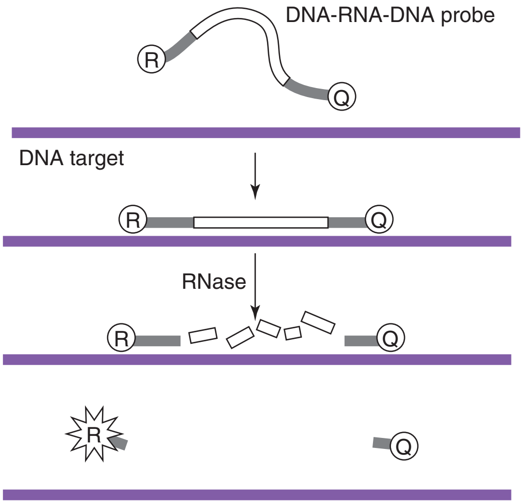


FIGURE 6.36 Cycling probe produces fluorescence only when the RNA probe binds to the DNA template. The RNA:DNA hybrid formed by the probe bound to the template is a substrate for RNase H, which digests the RNA probe and releases the reporter dye (R) from the vicinity of the quencher (Q).

STUDY QUESTIONS

1. A master mix of all components (except template) necessary for PCR contains what basic ingredients?

2. The final concentration of *Taq* polymerase is to be 0.01 units/µL in a 50-µL PCR. If the enzyme is supplied at 5 units/µL, how much enzyme would you add to the reaction?

   a. 1 µL
   b. 1 µL of a 1:10 dilution of *Taq*
   c. 5 µL of a 1:10 dilution of *Taq*
   d. 2 µL

3. Primer dimers result from

   a. high primer concentrations.
   b. low primer concentrations.
   c. high GC content in the primer sequences.
   d. 3′ complementarity in the primer sequences.

4. Which control is run to detect contamination?

   a. Negative control
   b. Positive control
   c. Molecular-weight marker
   d. Reagent blank

5. Nonspecific extra PCR products can result from

   a. mispriming.
   b. high annealing temperatures.
   c. high agarose gel concentrations.
   d. omission of $MgCl_2$ from the PCR.

6. Using which of the following is an appropriate way to avoid PCR contamination?

   a. High-fidelity polymerase
   b. Hot-start PCR
   c. A separate area for PCR setup
   d. Fewer PCR cycles

7. How many copies of a target are made after 30 cycles of PCR?

   a. 2×30
   b. 2^{30}
   c. 30^2
   d. 30/2

8. What are the three steps of a standard PCR cycle?

9. Which of the following is a method for purifying a PCR product?

   a. Treating with uracil N glycosylase
   b. Adding divalent cations
   c. Putting the reaction mix through a spin column
   d. Adding DEPC

10. In contrast to standard PCR, real-time PCR is

    a. quantitative.
    b. qualitative.
    c. labor-intensive.
    d. sensitive.

11. In real-time PCR, fluorescence is *not* generated by which of the following?

    a. FRET probes
    b. TaqMan probes
    c. SYBR green
    d. *Tth* polymerase

12. Prepare a table that compares PCR, LCR, bDNA, TMA, Qβ replicase, and hybrid capture with regard to the type of amplification, target nucleic acid, type of amplicon, and major enzyme(s) for each.

13. Examine the following sequence. (The complementary strand is not shown.) You are devising a test to detect a mutation at the underlined position.

    5′TATTTAGTTA TGGCCTATAC ACTATTTGTG
    AGCAAAGGTG ATCGTTTTCT GTTTGAGATT
    TTTATCTCTT GATTCTTCAA AAGCATTCTG
    AGAAGGTGAG ATAAGCCCTG AGTCTCAGCT
    ACCTAAGAAA AACCTGGATG TCACTGGCCA
    CTGAGGAGC TTTGTTTCAAC CAAGTCATGT
    GCATTTCCAC GTCAACAGAA TTGTTTATTG
    TGACAGTT**A**T ATCTGTTGTC CCTTTGACCT
    TGTTTCTTGA AGGTTTCCTC GTCCCTGGGC
    AATTCCGCAT TTAATTCATG GTATTCAGGA
    TTACATGCAT GTTTGGTTA AACCCATGAGA

TTCATTCAGT TAAAAATCCA
GATGGCGAAT3′

Design one set of primers (forward and reverse) to generate an amplicon containing the underlined base.

The primers should be 20 bases long.

The amplicon must be 100 to 150 bp in size.

The primers must have similar melting temperatures (T_m), +/– 2°C.

The primers should have no homology in the last three 3′ bases.

a. Write the primer sequences 5′→3′ as you would if you were to order them from the DNA synthesis facility.
b. Write the T_m for each primer that you have designed.

14. How does nested PCR differ from multiplex PCR?

15. What replaces heat denaturation in strand displacement amplification?

References

1. Mullis K. The unusual origin of the polymerase chain reaction. *Scientific American* 1990;262:56–61.
2. Mullis K. The polymerase chain reaction. Nobel Lecture: The Nobel Prize in Chemistry, 1993.
3. Saiki R, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 1985;230:1350–1354.
4. Guyer R, Koshland DE. The molecule of the year. *Science* 1989;46:1543–1546.
5. Lawyer F, Stoffel S, Saiki RK, Myambo K, Drummond R, Gelfand DH. Isolation, characterization, and expression in *Escherichia coli* of the DNA polymerase gene from *Thermus aquaticus*. *Journal of Biological Chemistry* 1989;264:6427–6437.
6. Son J, Cho B, Hong S, Lee SH, Hoxha O, Haack AJ, Lee LP. Ultrafast photonic PCR. *Light: Science & Applications* 2015;4:e280.
7. Obeid P, Christopoulos TK, Crabtree HJ, Backhouse CJ. Microfabricated device for DNA and RNA amplification by continuous-flow polymerase chain reaction and reverse transcription-polymerase chain reaction with cycle number selection. *Analytical Chemistry* 2003;75:288–295.
8. Shaw K, Docker PT, Yelland JV, Dyer CE, Greenman J, Greenway GM, Haswell SJ. Rapid PCR amplification using a microfluidic device with integrated microwave heating and air impingement cooling. *Lab Chip* 2010;10:1725–1728.
9. Cone R, Fairfax MR. Protocol for ultraviolet irradiation of surfaces to reduce PCR contamination. *Genome Research* 1993;3:s15–s17.
10. Jelden K, Gibbs SG, Smith PW, Hewlett AL, Iwen PC, Schmid KK, Lowe JJ. Ultraviolet (UV)-reflective paint with ultraviolet germicidal irradiation (UVGI) improves decontamination of nosocomial bacteria on hospital room surfaces. *Journal of Occupational and Environmental Hygiene* 2017;14(6):456–460.
11. Klaschik S, Lehmann LE, Raadts A, Hoeft A, Stuber F. Comparison of different decontamination methods for reagents to detect low concentrations of bacterial 16S DNA by real-time-PCR. *Molecular Biology* 2002;22:231–242.
12. Meier A, Persing DH, Finken M, Bottger EC. Elimination of contaminating DNA within polymerase chain reaction reagents: implications for a general approach to detection of uncultured pathogens. *Journal of Clinical Microbiology* 1993;31:646–652.
13. Fox J, Ait-Khaled M, Webster A, Emery VC. Eliminating PCR contamination: is UV irradiation the answer? *Journal of Virological Methods* 1991;33:375–382.
14. Cleaver J, Crowley E. UV damage, DNA repair and skin carcinogenesis. *Frontiers in Bioscience* 2002;7:1024–1043.
15. Korbie D, Mattick JS. Touchdown PCR for increased specificity and sensitivity in PCR amplification. *Nature Protocols* 2008;3:1452–1456.
16. Tzanakaki G, Tsopanomichalou M, Kesanopoulos K, Matzourani R, Sioumala M, Tabaki A, Kremastinou J. Simultaneous single-tube PCR assay for the detection of *Neisseria meningitidis, Haemophilus influenzae* type b and *Streptococcus pneumoniae*. *Clinical Microbiology and Infection* 2005;11:386–390.
17. McClure J, DeLongchamp JZ, Conly JM, Zhang K. A novel multiplex PCR assay for the detection of chlorhexidine/quaternary ammonium, mupirocin and methicillin resistance genes with simultaneous discrimination of *Staphylococcus aureus* from coagulase-negative *Staphylococci*. *Journal of Clinical Microbiology* 2017;55(6):1857–1864.
18. Bellau-Pujol S, Vabret A, Legrand L, Dina J, Gouarin S, Petitjean-Lecherbonnier J, Pozzetto B, Ginevra C, Freymuth F. Development of three multiplex RT-PCR assays for the detection of 12 respiratory RNA viruses. *Journal of Virological Methods* 2005;126:53–63.
19. Jones D, Kamel-Reid S, Bahler D, Dong H, Elenitoba-Johnson K, Press R, Quigley N, Rothberg P, Sabath D, Viswanatha D, Weck K, Zehnder J. Laboratory practice guidelines for detecting and reporting BCR-ABL drug resistance mutations in chronic myelogenous leukemia and acute lymphoblastic leukemia. *Journal of Molecular Diagnostics* 2009;11:4–16.
20. Casabianca A, Orlandi C, Fraternale A, Magnani M. A new one-step RT-PCR method for virus quantitation in murine AIDS. *Journal of Virological Methods* 2003;110:81–90.
21. Fontecha N, Nieto MC, Andía D, Cisterna R, Basaras M. RNA extraction method is crucial for human papillomavirus E6/E7 oncogenes detection. *Virology Journal* 2017;14.
22. Koopmans M, Monroe SS, Coffield LM, Zaki SR. Optimization of extraction and PCR amplification of RNA extracts from

paraffin-embedded tissue in different fixatives. *Journal of Virological Methods* 1993;43:189–204.

23. Poggio G, Rodriguez C, Cisterna D, Freire MC, Cello J. Nested PCR for rapid detection of mumps virus in cerebrospinal fluid from patients with neurological diseases. *Journal of Clinical Microbiology* 2000;38:274–278.

24. Gibbons C, Awad-El-Kariem FM. Nested PCR for the detection of *Cryptosporidium parvum*. *Parasitology Today* 1999;15:345.

25. Pinti M, Nasi M, Moretti L, Mussini C, Petrusca D, Esposito R, Cossarizza A. Quantitation of CD95 and CD95L mRNA expression in chronic and acute HIV-1 infection by competitive RT-PCR. *Annals of the New York Academy of Sciences* 2000;926:46–51.

26. Higuchi R, Dollinger G, Walsh PS, Griffith R. Simultaneous amplification and detection of specific DNA sequences. *Biotechnology* 1992;10:413–417.

27. Higuchi R, Fockler C, Dollinger G, Watson R. Kinetic PCR. Real time monitoring of DNA amplification reactions. *Biotechnology* 1993;11:1026–1030.

28. Morris T, Robertson B, Gallagher M. Rapid reverse transcription-PCR detection of hepatitis C virus RNA in serum by using the TaqMan fluorogenic detection system. *Journal of Clinical Microbiology* 1996;34:2933–2936.

29. Holland P, Abramson RD, Watson R, Gelfand DH. Detection of specific polymerase chain reaction product by utilizing the 5′ to 3′ exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proceedings of the National Academy of Sciences* 1991;88: 7276–7280.

30. Lee L, Connell CR, Bloch W. Allelic discrimination by nick-translation PCR with fluorogenic probes. *Nucleic Acids Research* 1993;21:3761–3766.

31. Livak K, Marmaro J, Flood S. *Guidelines for designing TaqMan fluorescent probes for 5′ nuclease assays*. Foster City, CA: PE Applied Biosystems, 1995.

32. Furutani S, Naruishi N, Hagihara Y, Nagai H. Development of an on-site rapid real-time polymerase chain reaction system and the characterization of suitable DNA polymerases for TaqMan probe technology. *Analytical and Bioanalytical Chemistry* 2016;408:5641–5649.

33. Vet J, Marras SA. Design and optimization of Molecular Beacon real-time polymerase chain reaction assays. *Methods in Molecular Biology* 2005;288:273–290.

34. Whitcombe D, Theaker J, Guy SP, Brown T, Little S. Detection of PCR products using self-probing amplicons and fluorescence. *Nature Biotechnology* 1999;7:804–807.

35. Nordgård O, Oltedal S, Janssen EA, Gilje B, Kørner H, Tjensvoll K, Smaaland R. Comparison of a PNA clamp PCR and an ARMS/Scorpion PCR assay for the detection of K-ras mutations. *Diagnostic Molecular Pathology* 2012;21:9–13.

36. Thelwell NS, Millington S, Solinas A, Booth J, Brown T. Mode of action and application of Scorpion primers to mutation detection. *Nucleic Acids Research* 2000;28:3752–3761.

37. Didenko V. DNA probes using fluorescence resonance energy transfer (FRET): designs and applications. *BioTechniques* 2001;31: 1106–1121.

38. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, Vandesompele J, Wittwer CT. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry* 2009;55:611–622.

39. Johnson G, Nour AA, Nolan T, Huggett J, Bustin S. Minimum information necessary for quantitative real-time PCR experiments. *Methods in Molecular Biology* 2014;1160:5–17.

40. Kwoh D, Davis GR, Whitfield KM, Chappelle HL, DiMichele LJ, Gingeras TR. Transcription-based amplification system and detection of amplified human immunodeficiency virus type 1 with a bead-based sandwich hybridization format. *Proceedings of the National Academy of Sciences* 1989;86:1173–1177.

41. Deiman B, van Aarle P, Sillekens P. Characteristics and applications of nucleic acid sequence-based amplification (NASBA). *Molecular Biotechnology* 2002;20:163–180.

42. Hawkins T, Detter JC, Richardson PM. Whole genome amplification—applications and advances. *Current Opinion in Biotechnology* 2002;13:65–67.

43. Dean F, Hosono S, Fang L, Wu X, Faruqi AF, Bray-Ward P, Sun Z, Zong Q, Du Y, Du J, Driscoll M, Song W, Kingsmore SF, Egholm M, Lasken RS. Comprehensive human genome amplification using multiple displacement amplification. *Proceedings of the National Academy of Sciences* 2002;9:5261–5266.

44. Williams R, Peisajovich SG, Miller OJ, Magdassi S, Tawfik DS, Griffiths AD. Amplification of complex gene libraries by emulsion PCR. *Nature* 2008;3:545–550.

45. Welsh J, McClelland M. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research* 1990;18: 7213–7218.

46. Perucho M, Welsh J, Peinado MA, Ionov Y, McClelland M. Fingerprinting of DNA and RNA by arbitrarily primed polymerase chain reaction: applications in cancer research. *Methods in Enzymology* 1995;254:275–290.

47. Barany F. The ligase chain reaction in a PCR world. *PCR Methods and Applications* 1991;1:5–16.

48. Sun Y, Lu X, Su F, Wang L, Liu C, Duan X, Li Z. Real-time fluorescence ligase chain reaction for sensitive detection of single nucleotide polymorphism based on fluorescence resonance energy transfer. *Biosensors & Bioelectronics* 2015;74:705–710.

49. Hu Y, Jia H, Wang Y, Cheng Y, Li Z. Sensitive quantification of messenger RNA with a real-time ligase chain reaction by using a ribonucleotide-modified DNA probe. *Chemical Communications* 2014;50:13093–13095.

50. Su F, Wang L, Sun Y, Liu C, Duan X, Li Z. Highly sensitive detection of CpG methylation in genomic DNA by AuNP-based colorimetric assay with ligase chain reaction. *Chemical Communications* 2015;51:3371–3374.

51. Walker G, Little MC, Nadeau JG, Shank DD. Isothermal in vitro amplification of DNA by a restriction enzyme/DNA polymerase system. *Proceedings of the National Academy of Sciences* 1992;89:392–396.

52. Ichiyama S, Ito Y, Sugiura F, Iinuma Y, Yamori S, Shimojima M, Hasegawa Y, Shimokata K, Nakashima N. Diagnostic value of the strand displacement amplification method compared to those of Roche Amplicor PCR and culture for detecting mycobacteria in sputum samples. *Journal of Clinical Microbiology* 1997;35:3802–805.

53. Blumenthal T, Carmichael GG. RNA replication: function and structure of QBeta-replicase. *Annual Review of Biochemistry* 1979;48:525–548.

54. Horn T, Chang CA, Urdea MS. Chemical synthesis and characterization of branched oligodeoxyribonucleotides (bDNA) for use as signal amplifiers in nucleic acid quantification assays. *Nucleic Acids Research* 1997;25:4842–4849.

55. Kern D, Collins M, Fultz T, Detmer J, Hamren S, Peterkin JJ, Sheridan P, Urdea M, White R, Yeghiazarian T, Todd J. An enhanced-sensitivity branched-DNA assay for quantification of human immunodeficiency virus type 1 RNA in plasma. *Journal of Clinical Microbiology* 1996;34:3196–3202.

56. Lisby G. Application of nucleic acid amplification in clinical microbiology. *Methods in Molecular Biology* 1998;92:1–29.

57. Barathan M, Gopal K, Mohamed R, Ellegård R, Saeidi A, Vadivelu J, Ansari AW, Rothan HA, Ravishankar Ram M, Zandi K, Chang LY, Vignesh R, Che KF, Kamarulzaman A, Velu V, Larsson M, Kamarul T, Shankar EM. Chronic hepatitis C virus infection triggers spontaneous differential expression of biosignatures associated with T cell exhaustion and apoptosis signaling in peripheral blood mononucleocytes. *Apoptosis* 2015;20:466–480.

58. Clavel C, Masure M, Levert M, Putaud I, Mangeonjean C, Lorenzato M, Nazeyrollas P, Gabriel R, Quereux C, Birembaut P. Human papillomavirus detection by the hybrid capture II assay: a reliable test to select women with normal cervical smears at risk for developing cervical lesions. *Diagnostic Molecular Pathology* 2000;9:145–150.

59. Farthing A, Masterson P, Mason WP, Vousden KH. Human papillomavirus detection by hybrid capture and its possible clinical use. *Journal of Clinical Pathology* 1994;47:649–652.

60. Hessner M, Budish MA, Friedman KD. Genotyping of factor V G1691A (Leiden) without the use of PCR by invasive cleavage of oligonucleotide probes. *Clinical Chemistry* 2000;46:1051–1956.

61. Oh H, Smith CL. Evolving methods for single nucleotide polymorphism detection: factor V Leiden mutation detection. *Journal of Clinical Laboratory Analysis* 2011;25:259–288.

62. Germer J, Majewski D, Yung B, Mitchell PS, Yao JD. Evaluation of the invader assay for genotyping hepatitis C virus. *Journal of Clinical Microbiology* 2006;44:318–323.

63. Schutzbank T, Jarvis C, Kahmann N, Lopez K, Weimer M, Yount A. Detection of high-risk papillomavirus DNA with commercial invader-technology-based analyte-specific reagents following automated extraction of DNA from cervical brushings in ThinPrep media. *Journal of Clinical Microbiology* 2007;45:4067–4069.

64. Cloney L, Marlowe C, Wong A, Chow R, Bryan R. Rapid detection of mecA in methicillin resistant *Staphylococcus aureus* using cycling probe technology. *Molecular and Cellular Probes* 1999;13:191–197.

65. Modrusan Z, Marlowe C, Wheeler D, Pirseyedi M, Bryan RN. Detection of vancomycin-resistant genes vanA and vanB by cycling probe technology. *Molecular and Cellular Probes* 1999;13:223–231.

66. Merlino J, Rose B, Harbour C. Rapid detection of non-multidrug-resistant and multidrug-resistant methicillin-resistant *Staphylococcus aureus* using cycling probe technology for the *mecA* gene. *Journal of Clinical Microbiology and Infectious Disease* 2003;22:322–323.

67. Ihira M, Yamaki A, Kato Y, Higashimoto Y, Kawamura Y, Yoshikawa T. Cycling probe-based real-time PCR for the detection of human herpesvirus 6A and B. *Journal of Medical Virology* 2016;88:1628–1635.

68. Muraosa Y, Toyotome T, Yahiro M, Watanabe A, Shikanai-Yasuda MA, Kamei K. Detection of *Histoplasma capsulatum* from clinical specimens by cycling probe-based real-time PCR and nested real-time PCR. *Medical Mycology* 2016;54:433–438.

Chapter 7

Chromosomal Structure and Chromosomal Mutations

Outline

CHROMOSOMAL STRUCTURE AND ANALYSIS
    Chromosomal Compaction and Histones
    Chromosome Morphology
    Visualizing Chromosomes
DETECTION OF GENOME AND CHROMOSOMAL MUTATIONS
    Karyotyping
    Fluorescence In Situ Hybridization
        Interphase FISH
        Metaphase FISH
COMPARATIVE GENOME HYBRIDIZATION

Objectives

7.1 Define *mutations* and *polymorphisms*.
7.2 Distinguish the three types of DNA mutations: genome, chromosomal, and gene.
7.3 Describe chromosomal compaction and the proteins involved in chromatin structure.
7.4 Diagram a human chromosome, and locate the centromere, telomere, q arm, p arm, and express ideogram locations.
7.5 Illustrate the different types of structural mutations that occur in chromosomes.
7.6 Describe how karyotypes are made and state the karyotype designation of a normal male and female.
7.7 Identify the chromosomal abnormality in a given karyotype.
7.8 Interpret results from interphase and metaphase FISH analyses.
7.9 Distinguish between the effects of balanced and unbalanced translocations on an individual and the individual's offspring.
7.10 Interpret the results of comparative genome hybridization showing an amplification or deletion.

The human genome is all of the genes found in a single individual. The human genome consists of 2.9 billion nucleotide base pairs of DNA organized into 23 chromosomes. As **diploid** organisms, humans inherit a **haploid** set of genes (23 chromosomes) from each parent, so humans have two copies of every gene (except for some on the X and Y chromosomes). Each chromosome is a double helix of DNA, ranging from 246 million nucleotide base pairs in length in chromosome 1 (the largest) to 48 million nucleotide base pairs in chromosome 21 (Table 7.1). Genetic information is carried on the chromosomes in the form of the order or **sequence** of nucleotides in the DNA helix. A **phenotype** is a trait or group of traits resulting from transcription and translation of these genes. The **genotype** is the DNA nucleotide sequence responsible for a phenotype.

Genotypic analysis is performed to confirm or predict phenotype. In the laboratory, some changes in chromosome structure and chromosome number can be observed microscopically. Mutations at the nucleotide-sequence level are detected using biochemical or molecular methods. Alterations of the DNA sequence may affect not only the phenotype of an individual but the progeny of that individual as well. The latter, heritable changes are the basis for prediction of the phenotype in the next generation. The probability of inheritance of a phenotypic trait can be estimated using the logical methods of Mendel's laws of genetics and statistics.

A transmissible (inheritable) change in the DNA sequence is a mutation or polymorphism. Although these terms are sometimes used interchangeably, they do have slightly different meanings based on population genetics. A DNA sequence change that is present in a relatively small proportion of a population is a **mutation.** The more general term *variant* may be used, particularly to describe inherited or somatic sequence alterations, reserving the term *mutation* for rarer, usually somatic changes, for example, changes found only in tumor tissue. A variant that is present in at least 1% to 2% of a population is considered a **polymorphism.** Both mutations and polymorphisms may or may not produce phenotypic differences.

Polymorphisms are casually considered mutations that do not severely affect phenotype; this is generally true because any negative effect on survival and reproduction limits the persistence of a genotype in a

TABLE 7.1 Approximate Sizes of Human Chromosomes in Base Pairs

Chromosome	Millions of Base Pairs
1	249
2	242
3	198
4	191
5	181
6	171
7	159
8	146
9	141
10	135
11	135
12	133
13	115
14	107
15	102
16	90
17	83
18	78
19	59
20	63
21	48
22	49
X	155
Y	59

population. Some polymorphisms, however, are maintained in a population through a balance of positive and negative phenotype. The classic example is sickle cell anemia, a condition caused by a single-base substitution in the gene that codes for hemoglobin. The alteration is regarded as a mutation, but it is really a **balanced polymorphism.** In addition to causing abnormal red blood cells, the genetic alteration results in resistance to infection by *Plasmodium* species, that is, resistance to malaria. The beneficial trait provides a survival and reproductive advantage that maintains the polymorphism in a relatively large proportion of affected populations. Examples of benign polymorphisms with no selective advantage are the ABO blood groups and the major histocompatibility complex and polymorphisms used for human identification and paternity testing.

DNA mutations can affect a single nucleotide or millions of nucleotides, even whole chromosomes, and thus can be classified into three categories: gene, chromosome, and genome mutations. Gene mutations affect single genes and are often, but not always, small changes in the DNA sequence. Chromosome mutations affect the structures of entire chromosomes. These changes require movement of large chromosomal regions (hundreds of thousands to millions of base pairs) either within the same chromosome or to another chromosome.

Genome mutations are changes in the number of chromosomes. A cell or cell population with a normal complement of chromosomes is **euploid.** Genome mutations result in cells that are **aneuploid.** Aneuploidy is mostly observed as increased numbers of chromosomes because the loss of whole chromosomes is generally not compatible with survival. Aneuploidy in diploid organisms can result when there are more than two copies of a single chromosome or when there are multiple copies of one or more chromosomes. Down syndrome is an example of a disease resulting from aneuploidy, where there are three copies, or trisomy, of chromosome 21.

Detection of mutations in the laboratory ranges from direct visualization of genome and chromosomal mutations under the microscope to molecular methods to detect single-base changes. Methods used for detection of genome and chromosomal mutations are discussed in this chapter.

CHROMOSOMAL STRUCTURE AND ANALYSIS

Chromosomal Compaction and Histones

An important concept in the understanding of chromosomes is that chromosome behavior is dependent on chromosome structure as well as DNA sequence.[1] Genes with identical DNA sequences will behave differently, depending on the chromosomal location or the surrounding nucleotide sequence of the gene. It is a well-known phenomenon that a gene inserted or moved into a different chromosomal location may be expressed (transcribed and translated) differently than it was in its original position. This is called the **position effect.** Furthermore, different sequences can have the same functional effect, such as the centromeres (where the chromosome attaches to the spindle apparatus for proper segregation during cell division), which are not defined by specific DNA sequences.[2]

A eukaryotic chromosome is a double helix of DNA. A cell nucleus contains 4 cm of double helix, which must be compacted, both to fit into the nucleus and to accurately segregate in **mitosis.** An extended DNA double helix undergoes an 8,000-fold **compaction** to make a metaphase chromosome (Fig. 7.1).[3]

The winding of DNA onto **histones,** the most abundant proteins in the cell, is the first step in compaction. Approximately 160 to 180 base pairs (bp) of DNA are wrapped around a set of eight histone proteins (two each of H2a, H2b, H3, and H4) to form a **nucleosome.** Nucleosomes are visible by electron microscopy as 100-Å beadlike structures that are separated by short (70–90 bp) strands of a free double helix or linker DNA (Fig. 7.2). This "bead-on-a-string" arrangement comprises the 10-nm.

The structure of metaphase chromosomes is maintained by more than just histones. Metaphase chromatin is one-third DNA, one-third histones, and one-third nonhistone proteins. Nonhistone protein complexes, termed condensin I and condensin II, maintain mitotic chromosome structure.[3]

Historical Highlights

Before 1943, histones were thought to contain genetic information. Their function was later determined to be structural. It is now known that

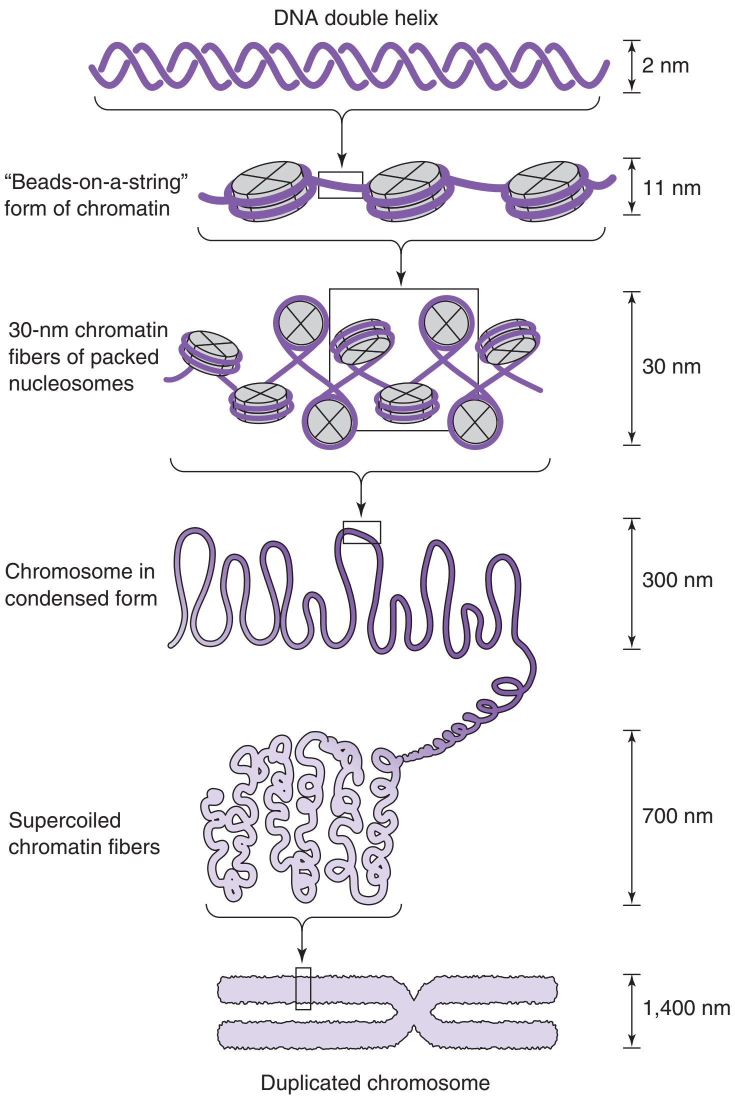

DNA double helix
2 nm
"Beads-on-a-string" form of chromatin
11 nm
30-nm chromatin fibers of packed nucleosomes
30 nm
Chromosome in condensed form
300 nm
Supercoiled chromatin fibers
700 nm
1,400 nm
Duplicated chromosome

FIGURE 7.1 DNA compaction into metaphase chromosomes. Histone wrapped in DNA forms the 10-nm chromatin strands found in transcriptionally active 10-nM DNA. Further compaction results in the closed 30-nm fibers found in transcriptionally silent DNA. (From Alberts B. *Molecular biology of the cell*. 4th ed. New York, NY: Garland Science, 2002.)

> in addition to their structural role, histones control access to and expression of DNA. Modification of histones, through acetylation, methylation, phosphorylation, or ubiquitination, alters DNA access and plays a role in other cellular functions, such as recombination, replication, and gene expression.

In the interphase nucleus, the 10-micron fiber is further coiled around histone H1 (or H5 in certain cells) into a thicker and shorter 30-nm or 30-micron fiber. The 30-nm interphase fibers represent the "resting state" of DNA. The fibers are locally relaxed into 10-nm fibers for DNA metabolism as required during the cell cycle. When the DNA is relaxed into 10-micron fibers for transcription or replication, the placement of nucleosomes along the double helix can be detected using nucleases (e.g., Mung bean nuclease, or DNase I). These enzymes cut the double helix in the part of the double helix that is exposed between the histones.

The 30-nm interphase fibers are looped onto protein scaffolds to form 300-nm fibers before entry into the M phase of the cell cycle (mitosis), and the looped fibers are wound into 700-nm solenoid coils.[4] The 700-nm

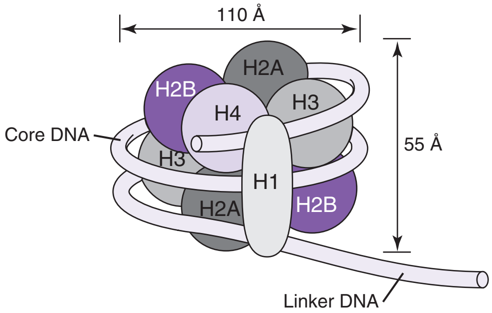


FIGURE 7.2 DNA wrapped around eight histone proteins (two each of histones 2A, 2B, 3, and 4) forms a nucleosome. A further association with histone H1 coils the nucleosomal DNA into a 30-nm fiber.

coils are compacted into the 1,400-nm fibers that can be seen microscopically in metaphase nuclei and as **karyotypes** in laboratory testing.

In the 30-nm interphase chromatin fiber, the internucleosomal DNA is wound into a solenoid coil. Loss of this level of organization is the first classic indicator of **apoptosis,** or programmed cell death. The 30-nm fibers are uncoiled, and the exposed linker DNA between the nucleosomes becomes susceptible to digestion by intracellular nucleases. The DNA wrapped into the nucleosomes remains intact so that DNA isolated from apoptotic cells contains "ladders," or discrete multiples of approximately 180 bp. These ladders resolved by agarose gel were one of the earlier molecular methods used to detect or confirm cell death by apoptosis (Fig. 7.3).

Chromosome topology (state of compaction of the DNA double helix) affects gene activity. Highly compacted DNA is less available for RNA transcription. The more highly compacted state of DNA is closed chromatin, or heterochromatin (in contrast to open chromatin, or euchromatin). Maintenance of heterochromatin throughout interphase may require condensin proteins or condensin-like protein complexes (nonhistone proteins).

Advanced Concepts

Members of a family of proteins called SMC proteins control chromosome condensation in eukaryotes and other aspects of chromosome behavior, including chromosome segregation in prokaryotes. Two of the SMC proteins, XCAP-C and XCAP-E, first isolated from frog eggs,[5] are integral parts of the condensin complex, a protein scaffold structure that can be isolated from both mitotic and interphase cells. In the presence of topoisomerase, this complex can wrap DNA around itself in a reaction driven by adenosine triphosphate (ATP). SMC family proteins also play a role in the repair of chromosomal breaks.[6] Although the exact role of the SMC proteins in chromosome condensation and decondensation is not yet completely defined, this ability to change chromosome architecture is a significant feature of DNA metabolism.[7]

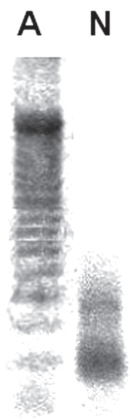


FIGURE 7.3 Apoptotic DNA (A) is characterized by the ladder seen on gel electrophoresis. The ladder forms as a result of nuclease activity on exposed linker regions of DNA. This is in contrast to randomly degraded DNA from necrotic cells (N).

Chromosome Morphology

Mitotic chromosomes have been distinguished historically by their relative size and centromere placement. The **centromere** is the site of attachment of the chromosome to the spindle apparatus. The connection is made between microtubules of the spindle and a protein complex, the **kinetochore,** that assembles at the centromere sequences (Fig. 7.4). At the nucleotide level, the centromere is composed of a set of highly repetitive **alpha satellite** sequences.[8,9] These repetitive sequences interfere with chromosome compaction so

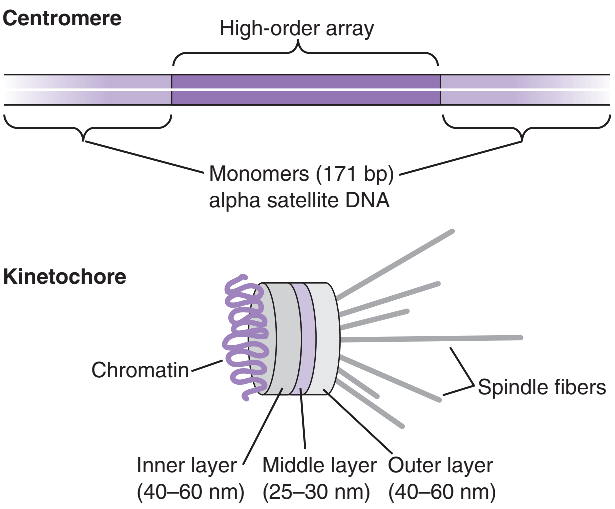


FIGURE 7.4 The centromere *(top)* consists of tandem repeats of 171 base-pair sequences flanking sets of single-repeat units, or monomers repeated in groups in a higher-order repeat array. The kinetochore *(bottom)* is a protein structure that connects the centromere chromatin to the spindle apparatus.

that microscopically, the centromere appears as a constriction in the metaphase chromosome. Chromosomes are **metacentric, submetacentric, acrocentric,** or **telocentric,** depending on the placement of the centromere (Fig. 7.5). The placement of the centromere divides the chromosome into arms. Metacentric chromosome arms are approximately equal in length, whereas one arm is longer than the other in submetacentric chromosomes. One arm is extremely small or missing in acrocentric or telocentric chromosomes, respectively.

Advanced Concepts

Some plants and insects have holocentric chromosomes. During cell division, these chromosomes associate with kinetochores along their entire length.

Human chromosomes are acrocentric or submetacentric and so have long and short arms (Table 7.2). The long arm of a chromosome is designated **q,** and the short arm is designated **p.** Acrocentric chromosomes have a ratio of long arm length:short arm length from 3:1 to 10:1. Chromosomes 13 to 15, 21, and 22 are considered acrocentric, but may be classified as subtelocentric.

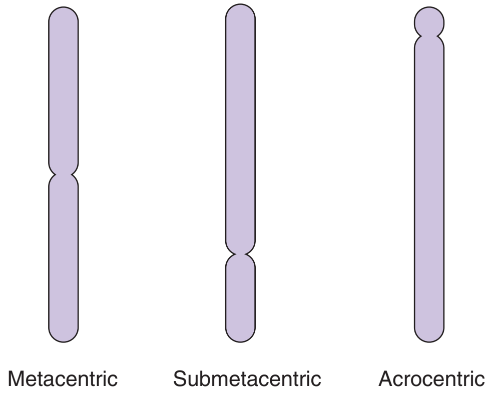


FIGURE 7.5 The arms of metacentric chromosomes *(left)* are of equal size. Submetacentric chromosomes *(center)* divide the chromosome into long arms and short arms. Acrocentric centromeres *(right)* are very near the ends of the chromosome.

Visualizing Chromosomes

Conventional cytological stains, such as Feulgen, Wright, and hematoxylin have been used to visualize chromosomes. An advance in the recognition of individual chromosomes was the discovery that fluorescent stains and chemical dyes can react with specific chromosome regions.[10] This region-specific staining forms reproducible patterns where portions of the chromosome accept or reject the stain. For cytogenetic analysis, this allows unequivocal identification of every chromosome and the direct detection of some chromosomal abnormalities. Underlying the region-specific staining is the implication that the reproducible staining patterns occur as a result of defined regional ultrastructures of the mitotic chromosomes.

When chromosomes are stained with the fluorescent dyes, quinacrine and quinacrine mustard, the resulting fluorescence pattern visualized after staining is **Q banding** (Fig. 7.6). This method was first demonstrated in 1970 by Caspersson, Zech, and Johansson.[11] The results of this work confirmed that each human chromosome could be identified by its characteristic banding pattern. Q banding gives a particularly intense staining of the human Y chromosome and thus may also be used to distinguish the Y chromosome in interphase

TABLE 7.2 Classification of Chromosomes by Size and Centromere Position

Group	Chromosomes	Description
A	1, 2	Large metacentric
	3	Large submetacentric
B	4, 5	Large submetacentric
C	6–12, X	Medium-sized submetacentric
D	13–15	Medium-sized acrocentric with satellites
E	16	Short metacentric
	17, 18	Short submetacentric
F	19, 20	Short metacentric
G	21, 22	Short acrocentric with satellites
	Y	Short acrocentric

nuclei. Because Q banding requires a fluorescent microscope, it is not as widely used as other stains that are detectable by light microscopy.

The chemical dye Giemsa stains in patterns, or **G bands,** similar to those seen in Q banding. The appearance of G banding differs, depending on the treatment of the chromosomes before staining. Mild treatment (2× standard saline citrate for 60 minutes at 60°C) yields the region-specific banding pattern comparable to that seen with fluorescent dyes.

Historical Highlights

Trypsin or other proteolytic extraction or denaturation of proteins before Giemsa staining was found to map structural aberrations more clearly and became the most commonly used staining method for analyzing chromosomes.[12,13] G bands were also produced by Feulgen staining after treatment with DNase I.[14] Harsher treatment of chromosomes (87°C for 10 min, then cooling to 70°C) before Giemsa staining will produce a pattern opposite to the G banding pattern called **R banding.** R bands can also be visualized after staining with acridine orange.[15] Alkali treatment of chromosomes results in centromere staining, or **C banding.**[16]

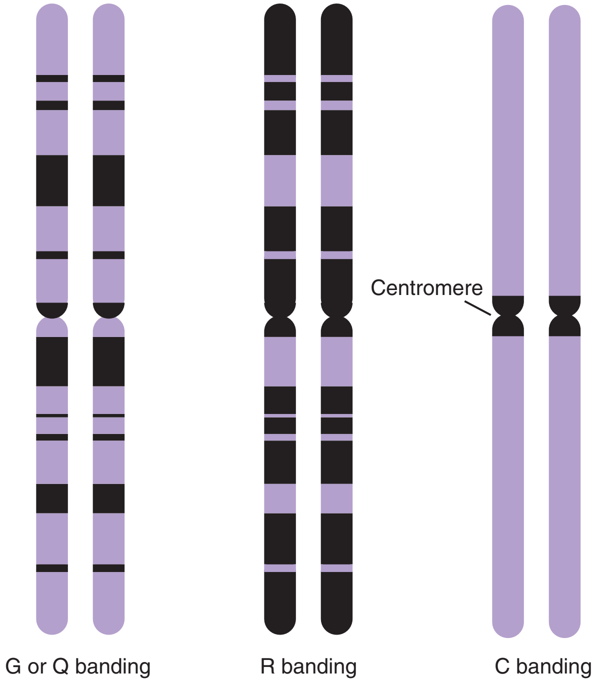


FIGURE 7.6 Reproducible staining patterns on chromosomes are used for identification and site location. Heterochromatin stains darkly by G or Q banding *(left)*; euchromatin stains darkly by R banding *(center)*; C banding stains centromeres *(right)*.

Centromere staining is absent in G-band patterns and may be associated with heterochromatin, the “quiet,” or poorly transcribed, sequences along the chromosomes that are also present around centromeres. In contrast, euchromatin, which is relatively rich in gene activity, may not be stained as much as heterochromatin in C banding. The correlation between heterochromatin and staining may also hold for noncentromeric G and Q bands. This association is complicated, however, because a variety of procedures and stains can produce identical banding patterns. The correlation of staining with heterochromatin is contradicted by observations of the X chromosome. Although one X chromosome is inactive and replicates later than the active X in females, both X chromosomes stain with equal pattern and intensity. Staining differences, therefore, must be due to other factors. Possible explanations for differential interactions with dye include differences in DNA compaction, sequences, and DNA-associated nonhistone proteins.

The number of visualized bands can be increased from about 300 to 500 per chromosome by staining chromosomes before they reach maximal metaphase condensation. This is called **high-resolution banding.**

Nucleolar organizing region (NOR) staining is another region-specific staining approach. Chromosomes treated with silver nitrate will stain specifically at the constricted regions, or stalks, on the acrocentric chromosomes.[16]

Staining of chromosomes with **4′,6-diamidino-2-phenylindole (DAPI)** was first described in 1976 as a way to detect mycoplasmal contamination in cell cultures.[17] DAPI binds to the surface grooves of double-stranded DNA and fluoresces blue under ultraviolet (UV) light (353-nm wavelength). DAPI is used to visualize chromosomes as well as whole nuclei.

Chromosome banding facilitates the detection of deletions, insertions, inversions, and other abnormalities and the identification of distinct chromosomal locations. For this purpose, the reproducible G-banding pattern has been ordered into regions, comprising bands and subbands. For example, in Figure 7.7, a site on the long arm (q) of chromosome 17 is located in region 1, band 1, sub-band 2, or 17q11.2.

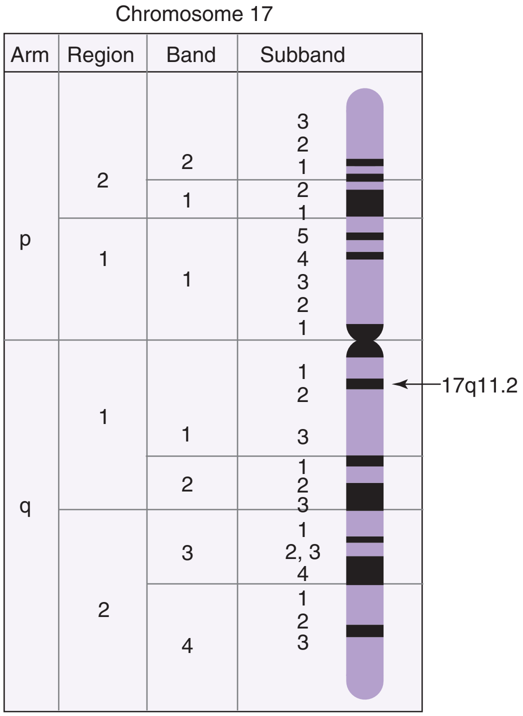


FIGURE 7.7 Identification of chromosomal location by G-band patterns. Locations are designated by the chromosome number 17 in this example, the arm q, the region 1, the band 1, and the sub-band 2.

DETECTION OF GENOME AND CHROMOSOMAL MUTATIONS

Karyotyping

Genome mutations, or aneuploidy, can be detected by indirect methods, such as flow cytometry, and more directly by karyotyping. A karyotype is the complete set of chromosomes in a cell. Karyotyping is the direct observation of metaphase chromosome structure by arranging metaphase chromosomes according to size. This requires collecting living cells and growing them in culture in the laboratory for 48 to 72 hours. Cell division is stimulated by addition of a **mitogen,** usually phytohemagglutinin. Dividing cells are then arrested in metaphase with Colcemid, an inhibitor of microtubule (mitotic spindle) formation. The chromosomes in dividing cells that arrest in metaphase will yield a **chromosome spread** when the cell nuclei are disrupted with hypotonic buffer. The 23 pairs of chromosomes can then be assembled into an organized display, or karyotype, according to their size and centromere placement (Fig. 7.8). Aneuploidy may be observed affecting several chromosomes[18] (Fig. 7.9) or a single chromosome (Fig. 7.10).

Karyotyping can also be used to detect chromosomal mutations such as **translocations,** which are the exchange of genetic material between chromosomes. Translocations can be of several types. In **reciprocal translocations,** parts of two chromosomes exchange; that is, each chromosome breaks, and the broken chromosomes reassociate or recombine with one another.

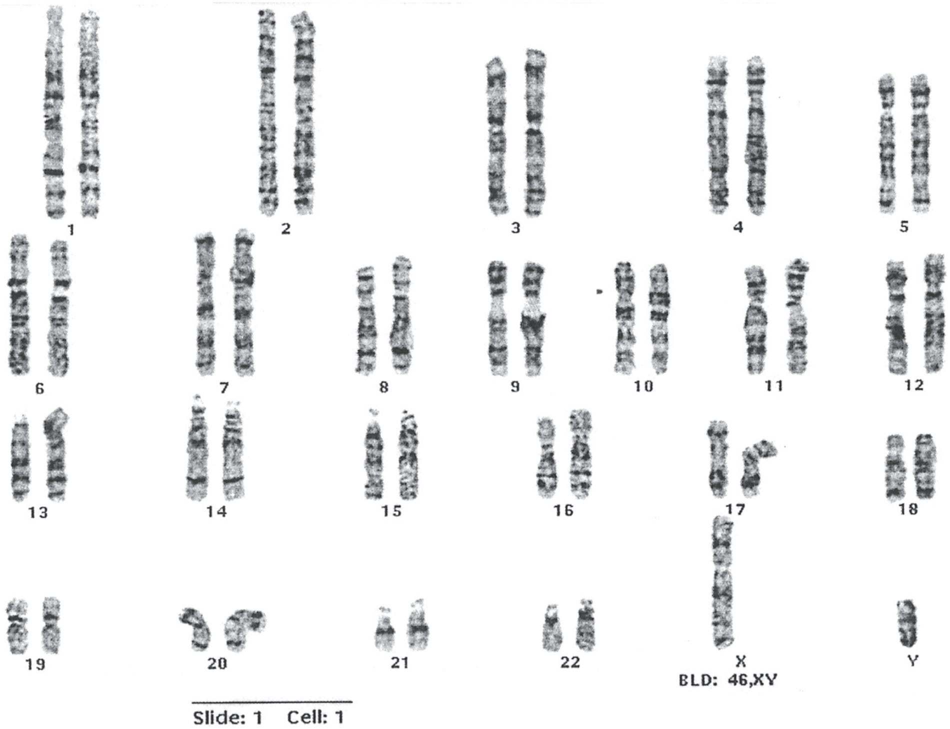


FIGURE 7.8 A normal male karyotype. There are 22 pairs of autosomes, one inherited from each parent, and one pair of sex chromosomes, XY. This karyotype is designated 46,XY.

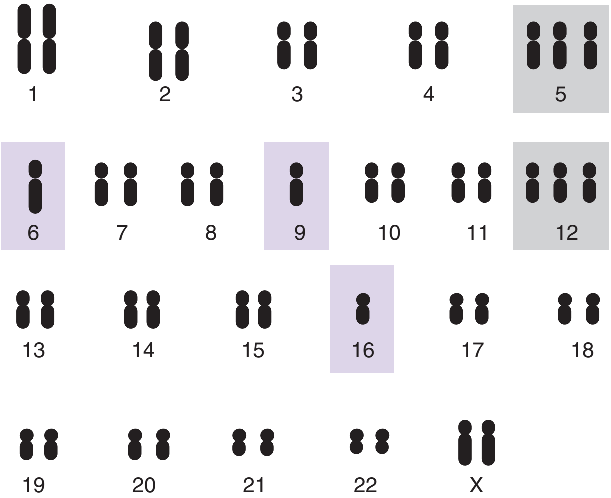


FIGURE 7.9 Aneuploidy involving multiple chromosomes. Chromosomes 5 and 12 are show trisomy; chromosomes 6, 9, and 16 are show monosomy.

When this type of translocation results in no gain nor loss of chromosomal material, it is **balanced** (Figs. 7.11 and 7.12). Balanced translocations may occur, therefore, without phenotypic effects. Balanced translocations in germ cells (cells that give rise to eggs or sperm) can, however, become unbalanced by not assorting properly during meiosis; as a result, they affect the phenotype of offspring. A robertsonian translocation involves the movement of the long arm of an acrocentric chromosome to the centromere of another acrocentric chromosome (Fig. 7.13). This type of translocation may also become unbalanced during reproduction, resulting in a net gain or loss of chromosomal material in the offspring.

Other types of chromosome mutations that are visible by karyotyping are shown in Figure 7.14. A **deletion** is a loss of chromosomal material. Large deletions covering millions of base pairs can be detected using karyotyping; smaller **microdeletions** are not always easily seen using this technique. An **insertion** is a gain of chromosomal

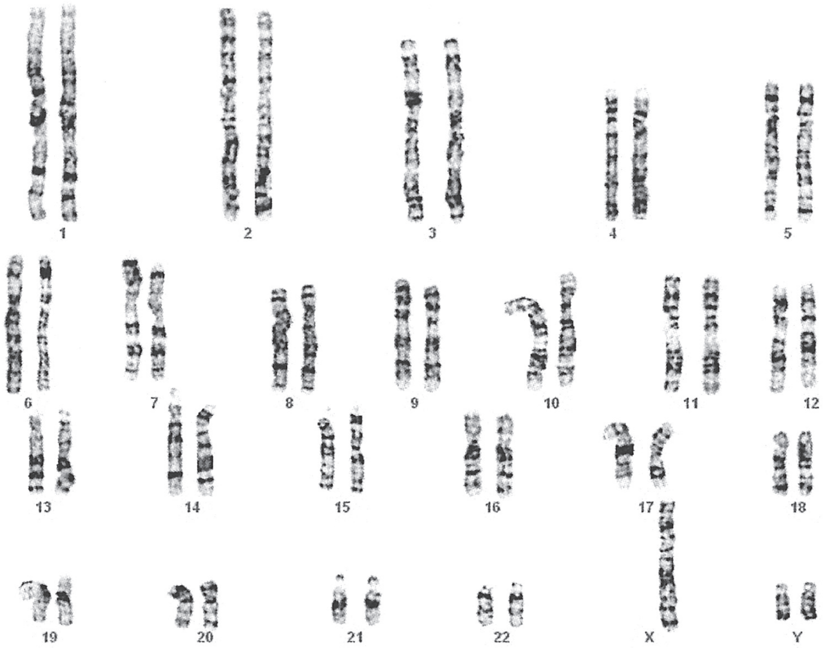


FIGURE 7.10 Aneuploidy involving Y chromosome disomy (XYY syndrome). This is designated 47,XYY.

material. The inserted sequences arise from duplication of particular regions within the affected chromosome or from fragments of other chromosomes. As with deletions, insertions cause altered banding patterns and a change in the size of the chromosomes. **Inversions** result from excision, flipping, and reconnecting chromosomal material within the same chromosome. **Pericentric inversions** include the centromere in the inverted region, whereas **paracentric inversions** involve sequences within one arm of the chromosome. An **isochromosome** is a metacentric chromosome that results from transverse splitting of the centromere during cell division. Transverse splitting causes two long arms or two short arms to separate into daughter cells instead of normal chromosomes with one long arm and one short arm. The arms of an isochromosome are therefore equal in length and genetically identical. A **ring chromosome** results from deletion of genetic regions from ends of the chromosome and a joining of the ends to form a ring. A **derivative chromosome** is an abnormal chromosome consisting of translocated or otherwise rearranged parts from two or more unidentified chromosomes joined to a normal chromosome.

Historical Highlights

The first chromosome mutations associated with human disease were visualized in the 1960s in leukemia cells. Peter Nowell and David Hungerford observed an abnormally small chromosome 22 in leukemia cells, which was labeled the "Philadelphia" chromosome. A few years later, Janet Rowley, using chromosome banding, noted that tumor cells not only lost genetic material, but they exchanged it. In 1972 she first described the translocation between chromosomes 8 and 21, t(8;21) in patients with acute myeloblastic leukemia. In that same year, she demonstrated that the Philadelphia chromosome was the result of a reciprocal exchange between chromosome 9 and chromosome 22.[19] She went on to identify additional reciprocal translocations in other diseases: the t(14;18) translocation in follicular lymphoma and the t(15;17) translocation in acute promyelocytic leukemia. This was also the first evidence that cancer had a genetic basis.

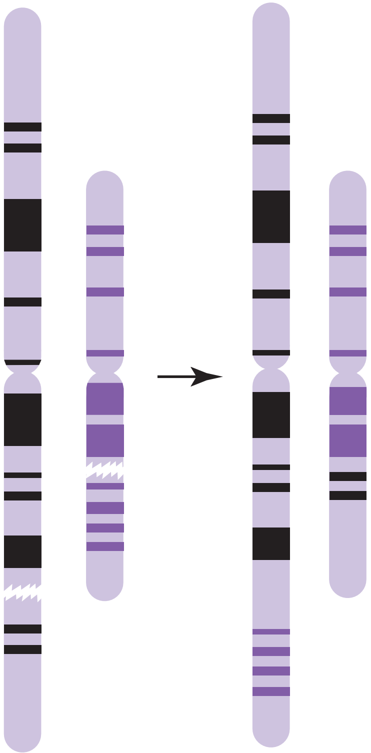


FIGURE 7.11 A balanced reciprocal translocation. At the resolution of karyotyping, no chromosomal material (banding) is lost.

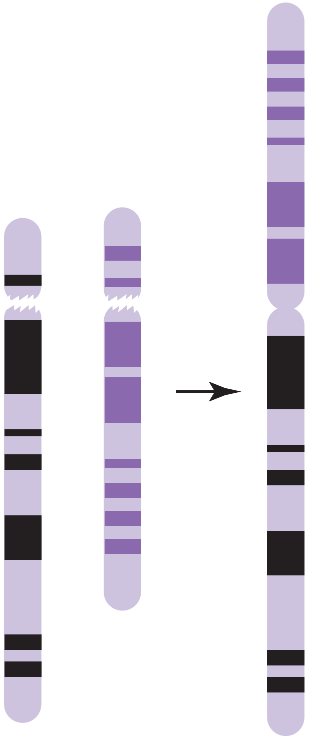


FIGURE 7.13 An isochromosome. The long arms of two acrocentric chromosomes are joined at the centromere.

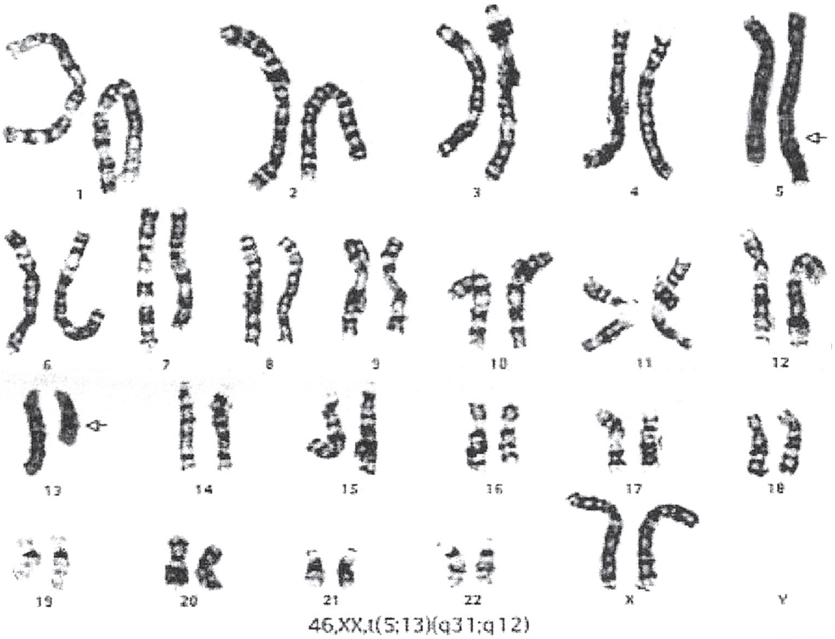

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Y
46,XX,t(5;13)(q31;q12)

FIGURE 7.12 A karyotype showing a balanced reciprocal translocation between chromosomes 5 and 13. This is designated 46,XX,t(5;13).

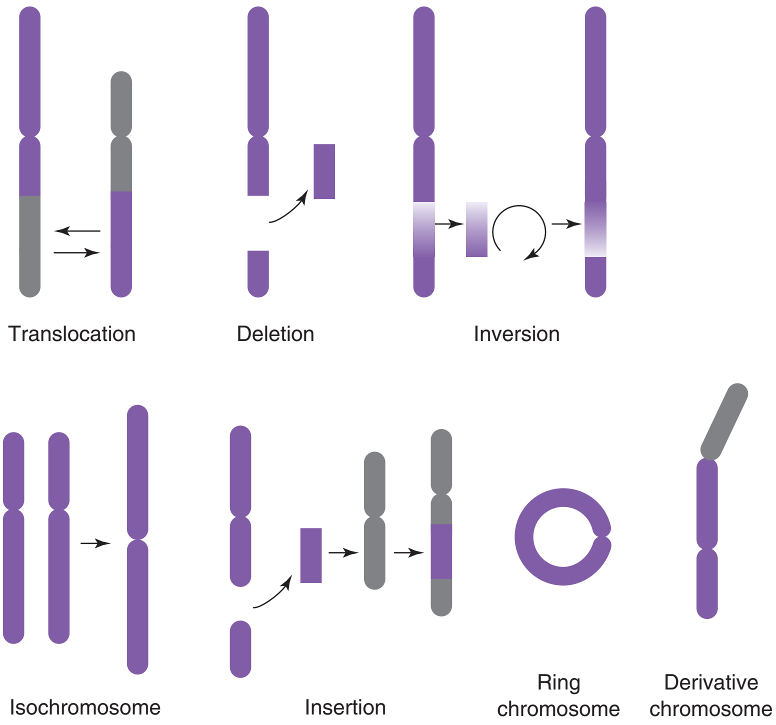


FIGURE 7.14 Chromosome mutations involving alterations in chromosome structure. In addition to translocations, ring and derivative chromosomes may or may not result in loss of chromosomal material. Insertions without duplication of the inserted regions and deletions will result in gain or loss of DNA.

Results of karyotyping analyses are expressed as the number of chromosomes per nucleus (normal is 46), the type of sex chromosomes (normal is XX or XY), followed by any genetic abnormalities observed. A normal karyotype is 46,XX in a female or 46,XY in a male. A karyotype showing 46,XX,del(7)(q13) denotes a deletion in the long arm q of chromosome 7 at region 1, band 3. A karyotype showing 46,XY,t(5;17)(p13.3;p13) denotes a translocation t between the short arms p of chromosomes 5 and 17 and region 1, band 3, subband 3, and region 1, band 3, respectively. A karyotype showing 47,XX+21 is the karyotype of a female with Down syndrome resulting from an extra chromosome 21. Klinefelter syndrome is caused by an extra X chromosome in males, for example, 47,XXY. Table 7.3 lists some of the terms used in expressing karyotypes.

Manual assembly of karyotypes from microscopic images has been replaced by software systems that electronically arrange the chromosomes from the image of the chromosome spread. Although this automation is highly applicable to the static images of chromosome spreads, such an automated system is more difficult to apply to fluorescent chromosome analysis (see following discussion), which would require a motorized scanning stage, automated area selection on the slide, and signal evaluation.

Fluorescence In Situ Hybridization

Fluorescence in situ hybridization (FISH) is a method widely used to detect protein and RNA as well as DNA structures in place in the cell, or in situ.[20] FISH offers a more rapid assay with higher resolution and flexibility than karyotyping.[21] FISH targets specific sequences of chromosomes with fluorescent probes. Even though FISH offers higher resolution than karyotyping for specific targets, it is limited to the regions complementary to the FISH probes. Probes are designed to hybridize to critical areas that are amplified, deleted, translocated, or otherwise rearranged in disease states. Unlike karyotyping that is performed under a light microscope, FISH analysis requires a fluorescence microscope that will excite fluorescent emission for the probes and special filters for detection of fluors emitting at different wavelengths.

TABLE 7.3 A List of Descriptive Abbreviations

Abbreviation	Indication
+	Gain
–	Loss
del	Deletion
der	Derivative chromosome
dup	Duplication
ins	Insertion
inv	Inversion
i, iso	Isochromosome
mat	Maternal origin
pat	Paternal origin
r	Ring chromosome
t	Translocation
tel	Telomere (end of chromosome arm)

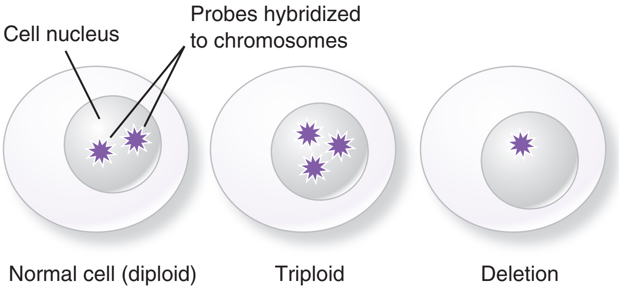


FIGURE 7.15 FISH analysis using centromeric probes for a normal diploid cell *(left),* triploidy or trisomy *(center),* and deletion or monosomy *(right).*

Analysis of signals from FISH also requires expertise in reading signals in three dimensions.

Interphase FISH

In contrast to karyotyping, **interphase FISH** does not require culturing of cells. Because growing cells in culture is not required, interphase FISH methods are used commonly to study prenatal samples, tumors, and hematological malignancies, not all of which are conveniently brought into metaphase in culture.

For FISH cytogenetic analysis, fixed cells are permeabilized and exposed to a probe. The probe is a 60- to 200-kb fragment of DNA attached covalently to a fluorescent molecule. The probe will hybridize, or bind, to its complementary sequences in the cellular DNA. In interphase FISH, the bound probe is visualized under a fluorescent microscope as a point of fluorescent light in the nucleus of the cell. Probes are designed to be complementary to a particular chromosome or chromosomal locus so that the image under the microscope will correlate with the state of that chromosome or locus. For example, a probe to any unique region on chromosome 22 should yield an image of two signals per nucleus, reflecting the two copies of chromosome 22 in the somatic cell nucleus (Fig. 7.15). A deletion or duplication of the DNA that is hybridized to the probe will result in a nucleus with only one signal or more than two signals, respectively. Multiple probes spanning large regions are used to detect regional deletions.[22]

Translocations or other rearrangements are detected using probes of different “colors” (or signals) complementary to regions on each chromosome taking part in the translocation. A normal nucleus will have two of each of the probe signals. A translocated chromosome will combine two of the probe signals, resulting in a loss of one of each signal in the nucleus. Analysis of translocation signals is sometimes complicated by false signals that result from two chromosomes landing close to one another in the nucleus, such that the bound probes give a signal similar to that exhibited by a translocation. These false signals may be distinguished from true translocations by the size of the fluorescent image, or by vertical focusing with the microscope. Accounting for false-positive signals as background noise limits the sensitivity of this assay.

The sensitivity of interphase FISH analysis for the detection of translocations is increased through the use of dual-color probes, or **dual-fusion probes.** These probes are mixtures of two single probes, each labeled with a different fluorescent dye. They are designed to bind to regions spanning the breakpoint of translocations. A translocation will be observed as a signal from both the translocation junction and the reciprocal of the

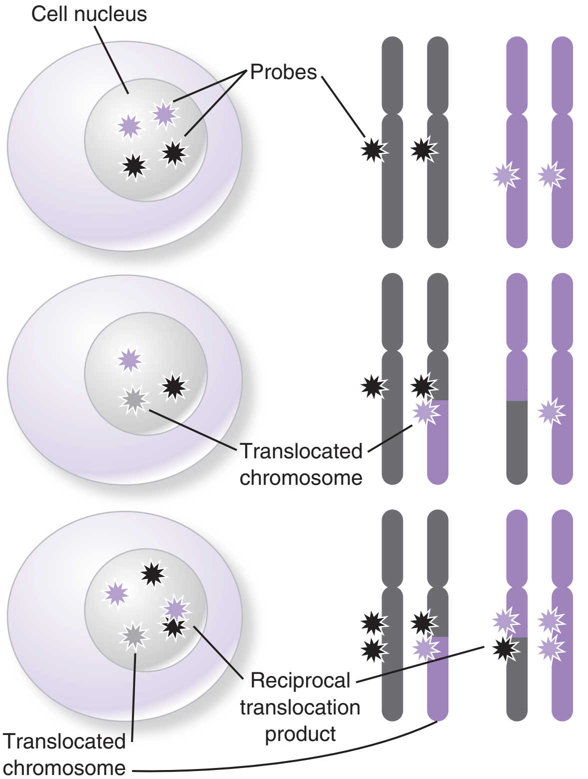


FIGURE 7.16 FISH analysis using distinct probes to detect a translocation. A normal nucleus has two signals from each probe *(top)*. A translocation involving the two chromosomes combines the two probe colors *(middle)*. Dual-fusion probes confirm the presence of the translocation by also giving a signal from the reciprocal breakpoint *(bottom)*. See Color Plate 2.

translocation junction, for example, t(9;22) and t(22;9) (Fig. 7.16). Dual-color **break-apart probes,** 0.6 to 1.5 Mb, are another approach to decrease background signals as well as to identify translocation events where one chromosome can recombine with multiple potential partners. These probes are designed to bind to the intact chromosome flanking the translocation breakpoint. When a translocation occurs, the two probes separate (Fig. 7.17). Sometimes called tri-FISH, break-apart probes are not the same as tricolor probes (see following discussion).

Centromeric probes (CEN probes) are designed to hybridize to highly repetitive alpha satellite sequences surrounding centromeres. These probes detect aneusomy of any chromosome. Combinations of centromeric probes and region-specific probes are often used to confirm deletions or amplifications in specific chromosomes. Addition of a CEN probe to dual-color probes comprises a tricolor probe and serves as a control for amplification or loss of one of the chromosomes involved in the translocation. For example, the IGH/MYC/CEP 8 Tri-color probes are a mixture of a 1.5-Mb–labeled probe, complementary to the immunoglobulin heavy-chain region (IGH) of chromosome 14; an approximately 750-kb distinctly labeled probe complementary to the *myc* gene on chromosome 8; and a CEN to chromosome 8.

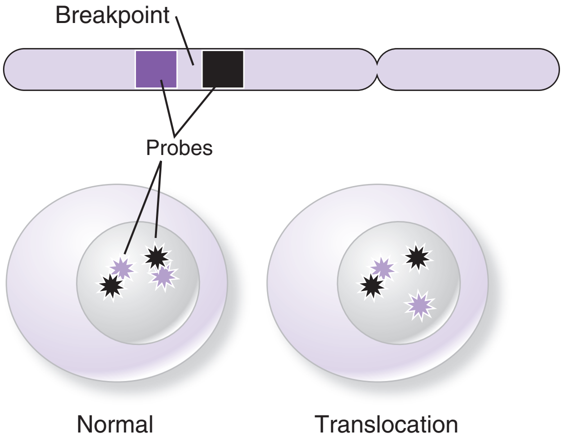


FIGURE 7.17 Break-apart probes bind to the chromosome flanking the translocation breakpoint region. Normal cells will display the combination signal *(bottom left),* and a translocation will separate the probe signals *(bottom right).*

Each chromosome **arm** has a unique set of repeat sequences located just before the end of the chromosome (the **telomere;** Fig. 7.18). These sequences have been studied to develop a set of DNA probes specific to the telomeres of all human chromosomes. **Telomeric probes** are useful for the detection of chromosome structural abnormalities, such as cryptic translocations or sub-telomeric deletions that are not easily visualized by standard karyotyping.

Because interphase cells for FISH do not require culturing of the cells and stimulating division, it is valuable for analysis of cells that do not divide well in culture, including fixed cells. Furthermore, because hundreds of cells are analyzed microscopically using FISH, the

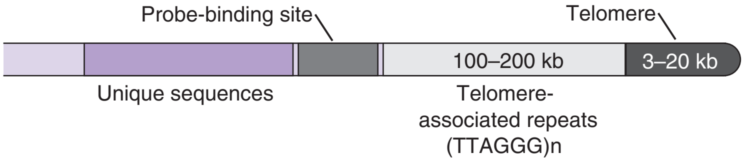


FIGURE 7.18 The binding sites for telomeric probes are unique sequences just next to the telomeric associated repeats and telomeric repeat sequences at the ends of chromosomes.

sensitivity of detection is higher than that of metaphase procedures, which commonly examine 20 spreads. A limitation of FISH, however, is the inability to identify chromosomal changes other than those at the specific binding region of the probe(s). In contrast, karyotyping, a more generic method, can detect any chromosomal change that causes changes in chromosomal size, number, or banding pattern within the sensitivity limits of the procedure.

Preparation of the sample is critical in interphase FISH analysis, both to permeabilize the cells for optimal probe–target interaction and to maintain cell morphology.[23] Optimal results are obtained if fresh interphase cells are incubated overnight (aging) after deposition on slides. After aging overnight, cells are treated with protease to minimize interference from cytoplasmic proteins and fixed with 1% formaldehyde to stabilize the nuclear morphology. Before DNA denaturation, the cells are dehydrated in graded concentrations of ethanol. Paraffin-embedded fixed tissues are dewaxed in xylene before protease and formaldehyde treatment.

The quality of the probe also has to be checked and its performance validated before use. Fluorescent probes (DNA with covalently attached fluorescent dyes) are usually purchased from vendors, which may also supply compatible hybridization reagents and controls. Nevertheless, the probe performance should be observed on control tissue before use on patient samples. Under a fluorescent microscope with the appropriate color-distinction filters, the signal from the probe should be bright, specific to the target in the cell nuclei, and free of high background noise. Probes differ in their signal characteristics and intensities; the technologist should become familiar with what to expect from a given probe on different types of tissues. This is facilitated by comparison with controls accompanying each run of samples.

As in Southern and northern blotting procedures, both probe and target must be denatured prior to hybridization. The amount of time taken to hybridize and use Cot-1 DNA (placental DNA enriched for repetitive sequences to reduce nonspecific binding) or facilitators such as dextran sulfate (to increase the effective probe concentration) depends on the sequence complexity of the probe. One to 10 micrograms of probe may be used in a hybridization volume of 3 to 10 µL. The hybridization of the probe on the target cells is performed at 37°C to 42°C in a humidified chamber. The slides are cover-slipped and sealed to optimize the hybridization conditions.

Following hybridization and rinsing off of the unbound probe, the sample is observed microscopically. The probe signals should be visible from entire intact nuclei. Adequate numbers of cells must be visible, but crowded cells where the nuclei and signals overlap do not yield accurate results. Furthermore, different tissue types have different image qualities and characteristics that must also be taken into account when assessing the FISH image. Another complication of FISH analysis is **photobleaching** (fading) or loss of probe signal emission (10^5 photons/second) due to photochemical destruction of the fluorophore molecules. For this reason, FISH slides should not be subjected to prolonged light exposure.

Metaphase FISH

Metaphase analysis has been enhanced by the development of fluorescent probes that bind to metaphase chromosomal regions or to whole chromosomes. **Metaphase FISH** allows analysis of small regions not visible by regular chromosome banding. Probes that cover the entire chromosome, or **whole chromosome paints,** are valuable for detecting these small or complex rearrangements (Fig. 7.19). By mixing combinations of five fluors and using special imaging software, **spectral karyotyping** can distinguish all 23 chromosomes by chromosome-specific colors.[24] This type of analysis can detect abnormalities that affect multiple chromosomes, as is sometimes found in cancer cells or immortalized cell lines.[25,26] Telomeric and centromeric probes are also

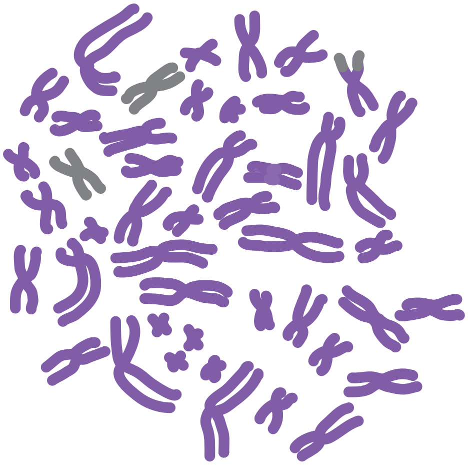


FIGURE 7.19 Chromosome painting showing a derivative chromosome formed by movement of a fragment of chromosome 12 (black) to an unidentified chromosome. See Color Plate 3.

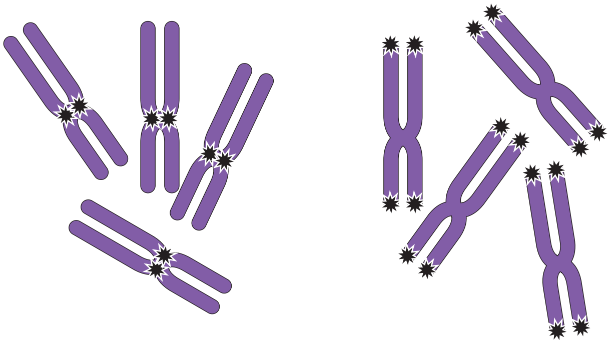


FIGURE 7.20 Centromeric *(left)* and telomeric *(right)* probes on metaphase chromosomes.

applied to metaphase chromosomes (Fig. 7.20) to detect aneuploidy and other genomic mutations.

Preparation of chromosomes for metaphase FISH procedures begins with the culture of cells for 72 hours. About 45 minutes before harvesting, colcemid is added to the cultures to arrest dividing cells in metaphase. The cells are then suspended in a hypotonic medium (0.075 M KCl) and fixed with methanol/acetic acid (3:1). The fixed-cell suspension is applied to an inclined slide and allowed to dry. A second treatment with 70% acetic acid may improve the chromosome spreading and decrease background noise. Condensed chromosome spreads, especially those from cultured metaphases, may be affected by temperature and humidity. Under a phase contrast microscope, the chromosomes should appear well separated with sharp borders. Cytoplasm should not be visible. Once the slide is dried, hybridization proceeds as discussed previously for interphase FISH.

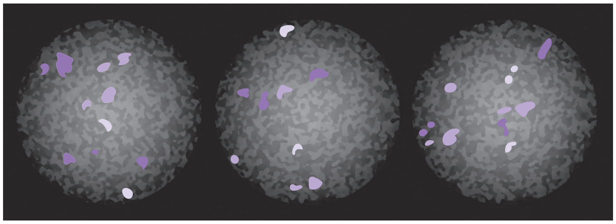


FIGURE 7.21 Multicolor FISH analysis simultaneously reveals structural or numerical abnormalities in three loci. See Color Plate 4.

Combinations of different locus-specific probes and chromosome paints can be used simultaneously, yielding more information than carrying out separate experiments on multiple specimens. The generic term for this is multicolor FISH (QMFISH or M-FISH). Imaging with 10 to 20 different probes or with a combination of region-specific probes and spectral karyotyping differentiates multiple chromosomes by spectral properties (Fig. 7.21). Simultaneously, M-FISH identifies specific chromosomal regions based on the presence or absence of the probe color visualized with specific filters. Analysis by both may show cryptic translocations and insertions as well as the chromosomal origins of marker chromosomes.[27]

Advanced Concepts

Quantitative FISH (Q-FISH; Fig. 7.22) is the analysis of repeated sequences by assessing the relative intensity of probe signals. Microscopic images are digitized on a charge-coupled device (CCD).[28,29] The signals are then measured by imaging software, quantifying FISH signals from each digital image. The relative intensity is compared between signals. The telomere Q-FISH technique has been applied to studies on telomere length.[30]

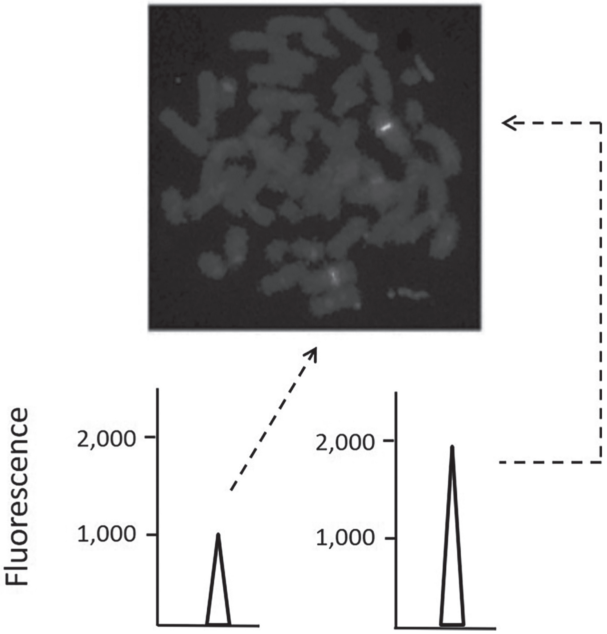


FIGURE 7.22 Fluorescent signals are digitized and quantified in quantitative FISH. The relative intensity of signals in multiple loci will demonstrate copy-number variations.

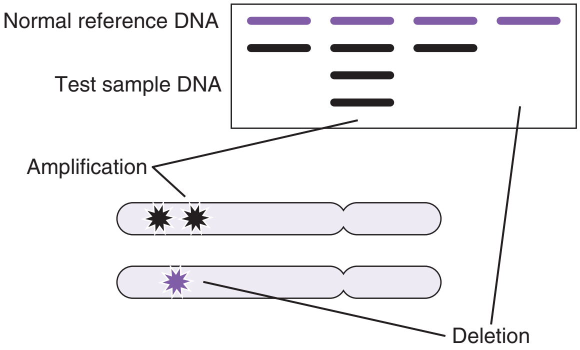


FIGURE 7.23 In CGH, the test sample is compared with a normal reference sample on a metaphase spread. Normally, test and reference signals are equal. A higher test signal denotes an amplification, and a higher reference signal denotes a deletion.

COMPARATIVE GENOME HYBRIDIZATION (CGH)

Intrachromosomal amplifications or deletions can be detected by **comparative genome hybridization (CGH).**[31] In this method, DNA from test and reference samples is labeled and used as a probe on a normal metaphase chromosome spread (Fig. 7.23). One advantage of CGH is its capability to identify the location of deletions or amplifications throughout the genome (Fig. 7.24); the resolution (precise identification of the amplified or deleted region), however, is not as high as can be achieved with array CGH.[32,33]

For CGH, the test DNA is isolated and labeled along with a reference DNA. Two colorimetrically distinct cyanine dyes, commonly Cy3 and Cy5, are used as fluorescent labels for the test and reference DNA. Cy3, which fluoresces at a wavelength of 550 nm, is often represented as "green," and Cy5, which fluoresces in the far-red region of the spectrum (650 to 667 nm), is represented as "red." Derivatives of these dyes, such as Cy3.5, which fluoresces in the red-orange region, are also available. Because these dyes fluoresce brightly and are water-soluble, they have been used extensively for CGH using imaging equipment.

Labeling (attachment of Cy3 or Cy5 dye to the test and reference DNA) is achieved by nick translation or primer extension in which nucleotides covalently attached to the dye molecules are incorporated into the DNA sequences. The dye nucleotides commonly used for this type of labeling are 5-amino-propargyl-2′-deoxycytidine 5′-triphosphate coupled to the Cy3 or Cy5 fluorescent dye (Cy3-AP3-dCTP, Cy5-AP3-dCTP) or 5-amino-propargyl-2′-deoxyuridine 5′-triphosphate coupled to the Cy3 or Cy5 fluorescent dye (Cy3-AP3-dUTP, Cy5-AP3-dUTP). Separate aliquots of test and reference DNA are labeled with different Cy3 and Cy5 dyes, respectively, before application to a normal metaphase spread. Test DNA is partially digested with DNase to produce fragments that will bind efficiently to the denatured DNA in a metaphase chromosome spread. Although CGH requires advanced technical expertise, it has shown value in identifying recurrent genomic imbalances not detected by karyotyping. Future advances and incorporation of high-density arrays will increase the ease and scope of use of this technology.

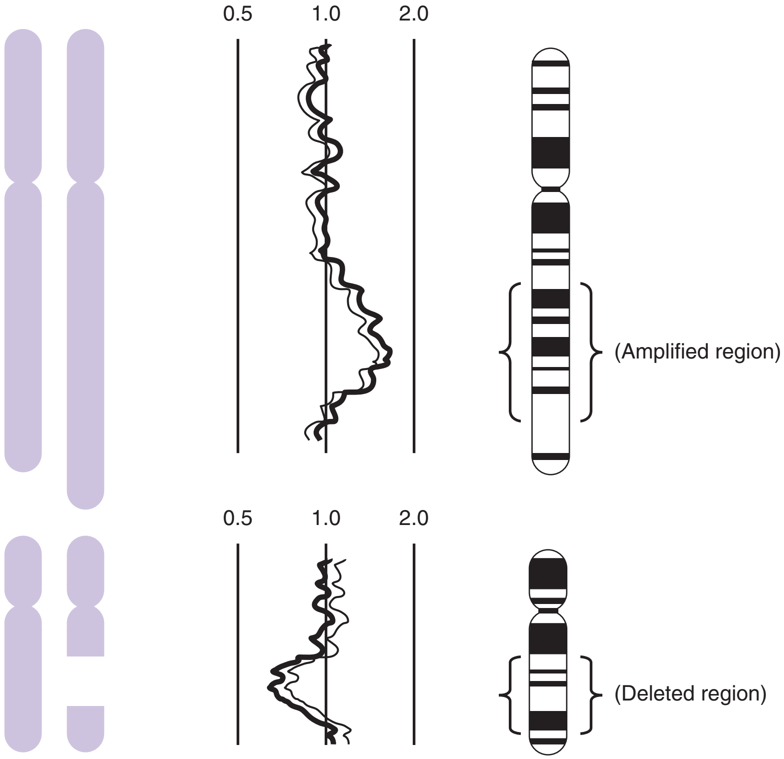


FIGURE 7.24 CGH analysis of four chromosomes of cells from a cancer cell line. Amplified or deleted areas are observed where the test and reference signals are not equal. The vertical lines in the center of the diagram represent the ratio of test/reference signals on the chromosomal spread showing excess test or excess reference signal *(right idiograms)*.

STUDY QUESTIONS

1. What chromosomal location is indicated by 15q21.1?

2. During interphase FISH analysis for the t(9;22) translocation, one nucleus was observed with two normal signals (one red for chromosome 22 and one green for chromosome 9) and one composite red/green signal. Five hundred other nuclei were normal. What is one explanation for this observation?

3. Is 47;XYY a normal karyotype?

4. Write the numerical and structural chromosomal abnormalities represented by the following genotypes:

   47,XY, +18
   46,XY, del(16)p(14)
   iso(Xq)
   46,XX del(22)q(11.2)
   45,X

5. A chromosome with a centromere located such that one arm of the chromosome is longer than the other arm is called

   a. metacentric.
   b. paracentric.
   c. telocentric.
   d. submetacentric.

6. A small portion from the end of chromosome 2 has been found on the end of chromosome 15, replacing the end of chromosome 15, which has

moved to the end of chromosome 2. This mutation is called a(n)

- **a.** reciprocal translocation.
- **b.** inversion.
- **c.** deletion.
- **d.** robertsonian translocation.

7. Phytohemagglutinin is added to a cell culture when preparing cells for karyotyping. The purpose of the phytohemagglutinin treatment is to

- **a.** arrest the cell in metaphase.
- **b.** spread out the chromosomes.
- **c.** fix the chromosomes on the slide.
- **d.** stimulate mitosis in the cells.

8. A centromeric probe is used to visualize chromosome 21. Three fluorescent signals are observed in the cell nuclei when stained with this probe. These results would be interpreted as consistent with

- **a.** a normal karyotype.
- **b.** Down syndrome.
- **c.** Klinefelter syndrome.
- **d.** technical error.

9. Cells were harvested from a patient's blood, cultured to obtain chromosomes in metaphase, fixed onto a slide, treated with trypsin, and then stained with Giemsa. The resulting banding pattern is called

- **a.** G banding.
- **b.** Q banding.
- **c.** R banding.
- **d.** C banding.

10. A FISH test with a centromere 13 probe is ordered for a suspected case of Patau syndrome (trisomy 13). How many signals per nucleus will result if the test is positive for Patau syndrome?

11. What would be the results if a centromere 13 probe was used on a case of Edward syndrome (trisomy 18)?

12. Angelman syndrome is caused by a microdeletion in chromosome 15. Which method, karyotyping or metaphase FISH, is better for accurate detection of this abnormality? Why?

13. The results of a CGH analysis of Cy3 (green)-labeled test DNA with Cy5 (red)-labeled reference DNA on a normal chromosome spread revealed a bright red signal along the short arm of chromosome 3. How is this interpreted?

- **a.** 3p deletion
- **b.** 3q deletion
- **c.** 3p amplification
- **d.** 3q amplification

14. A break-apart probe is used to detect a translocation. The results of FISH analysis show two signals in 70% of the nuclei counted and three signals in 30% of the nuclei. Is there a translocation present?

15. What FISH technique is most useful for the detection of multiple complex genomic mutations?

References

1. Murray A. How to compact DNA. *Science* 1998;282:425–427.
2. Black B, Foltz DR, Chakravarthy S, Luger K, Woods VL, Cleveland DW. Structural determinants for generating centromeric chromatin. *Nature* 2004;430:578–582.
3. Piskadlo E, Oliveira RA. Novel insights into mitotic chromosome condensation. *F1000Research* 2016;5:25.
4. Porter I, Khoudoli GA, Swedlow JR. Chromosome condensation: DNA compaction in real time. *Current Biology* 2004;14:R554–R556.
5. Hirano T, Mitchison TJ. A heterodimeric coiled-coil protein required for mitotic chromosome condensation in vitro. *Cell* 1994;79:449–458.
6. Kinoshita E, van der Linden E, Sanchez H, Wyman C. RAD50, an SMC family member with multiple roles in DNA break repair: how does ATP affect function? *Chromosome Research* 2009;17:277–288.
7. Hudson D, Ohta S, Freisinger T, Macisaac F, Sennels L, Alves F, Lai F, Kerr A, Rappsilber J, Earnshaw WC. Molecular and genetic analysis of condensin function in vertebrate cells. *Molecular and Cellular Biology* 2008;19:3070–3079.
8. Sullivan L, Chew K, Sullivan BA. Alpha satellite DNA variation and function of the human centromere. *Nucleus* 2017;13:331–339.
9. Waye J, Willard HF. Chromosome-specific alpha satellite DNA: nucleotide sequence analysis of the 2.0 kilobasepair repeat from the human X chromosome. *Nucleic Acids Research* 1985;13:2731–2743.

10. Huang H, Chen J. Chromosome bandings. *Methods in Molecular Biology* 2017;1541:59–66.

11. Caspersson T, Zech L, Johansson C. Differential banding of alkylating fluorochromes in human chromosomes. *Experimental Cell Research* 1970;60:315–319.

12. Seabright M. A rapid banding technique for human chromosomes. *Lancet* 1971;2:971–972.

13. Seabright M. The use of proteolytic enzymes for the mapping of structural rearrangements in the chromosomes of man. *Chromosoma* 1972;36:204–210.

14. Burkholder G, Weaver M. DNA protein interactions and chromosome binding. *Experimental Cell Research* 1977;110:251–262.

15. Bobrow M, Madan, K. The effects of various banding procedures on human chromosomes studied with acridine orange. *Cytogenetics and Cell Genetics* 1973;12:145–156.

16. Arrighi F, Hsu TC. Localization of heterochromatin in human chromosomes. *Cytogenetics* 1971;10:81–86.

17. Jagielski M, Zaleska M, Kaluzewski S, Polna I. Applicability of DAPI for the detection of mycoplasma in cell cultures. *Medycyna doświadczalna i mikrobiologia* 1976;28:161–173.

18. Grimm D. Genetics. Disease backs cancer origin theory. *Science* 2004;306:389.

19. Mayall B, Carrano AV, Moore DH, Rowley JD. Quantification by DNA-based cytophotometry of the 9q+/22q-chromosomal translocation associated with chronic myelogenous leukemia. *Cancer Research* 1977;37:3590–3593.

20. Hu L, Ru K, Zhang L, Huang Y, Zhu X, Liu H, Zetterberg A, Cheng T, Miao W. Fluorescence in situ hybridization (FISH): an increasingly demanded tool for biomarker research and personalized medicine. *Biomarker Research* 2014;2:1–13.

21. Wan T. Cytogenetics: an introduction. *Methods in Molecular Biology* 2017;1541:1–10.

22. John S, Erming T, Jeffrey S, Klaus B, Reid L, Ashraf B, Christopher M, Guido Tri, Joshua E, Bart B. High incidence of chromosome 13 deletion in multiple myeloma detected by multiprobe interphase FISH. *Blood* 2000;96:1505–1511.

23. Van Stedum S, King W. Basic FISH techniques and troubleshooting. *Methods in Molecular Biology* 2002;204:51–63.

24. Macville M, Veldman T, Padilla-Nash H, Wangsa D, O'Brien P, Schrack E, Ried T. Spectral karyotyping, a 24-colour FISH technique for the identification of chromosomal rearrangements. *Histochemistry and Cell Biology* 1997;108:299–305.

25. Liang J, Ning Y, Wang R, Padilla-Nash HM, Schrock E, Soenksen D, Nagarajan L, Ried T. Spectral karyotypic study of the HL-60 cell line: detection of complex rearrangements involving chromosomes 5, 7, and 16 and delineation of critical region of deletion on 5q31.1. *Cancer Genetics and Cytogenetics* 1999;113:105–109.

26. Mehra S, Messner H, Minden M, Chaganti RS. Molecular cytogenetic characterization of non-Hodgkin lymphoma cell lines. *Genes Chromosomes Cancer* 2002;33:225–234.

27. Bayani J, Squire J. Multi-color FISH techniques. *Current Protocols in Cell Biology* 2004;22.

28. Iourov I, Soloviev IV, Vorsanova SG, Monakhov VV, Yurov YB. An approach for quantitative assessment of fluorescence in situ hybridization (FISH) signals for applied human molecular cytogenetics. *Journal of Histochemistry and Cytochemistry* 2005;53:401–408.

29. Iourov I. Quantitative fluorescence in situ hybridization (QFISH). *Methods in Molecular Biology* 2017;1541:143–149.

30. Ikeda H, Aida J, Hatamochi A, Hamasaki Y, Izumiyama-Shimomura N, Nakamura K, Ishikawa N, Poon SS, Fujiwara M, Tomita K, Hiraishi N, Kuroiwa M, Matsuura M, Sanada Y, Kawano Y, Arai T, Takubo K. Quantitative fluorescence in situ hybridization measurement of telomere length in skin with/without sun exposure or actinic keratosis. *Human Pathology* 2014;45:473–480.

31. Lapierre J, Cacheux V, Da Silva F, Collot N, Hervy N, Wiss J, Tachdjian G. Comparative genomic hybridization: technical development and cytogenetic aspects for routine use in clinical laboratories. *Annales de Genetique* 1998;41:56–62.

32. Zhang C, Cerveira E, Romanovitch M, Zhu Q. Array-based comparative genomic hybridization (aCGH). *Methods in Molecular Biology* 2017;1541:167–179.

33. Haraksingh R, Abyzov A, Urban AE. Comprehensive performance comparison of high-resolution array platforms for genome-wide copy number variation (CNV) analysis in humans. *BMC Genomics* 2017;18:1–14.

Chapter 8

Gene Mutations

Outline

TYPES OF GENE MUTATIONS
DETECTION OF GENE MUTATIONS
- *Biochemical Methods*
    - Enzyme Immunoassays
    - Immunohistochemistry
    - High-Performance Liquid Chromatography (HPLC)
    - Gas Chromatography
    - Mass Spectrometry
- *Nucleic Acid Analyses*
    - Hybridization-Based Methods
    - Sequencing (Polymerization)-Based Methods
    - Enzymatic and Chemical Cleavage Methods
    - Other Methods

GENE VARIANT NOMENCLATURE
GENE NAMES

Objectives

8.1 Compare the phenotypic consequences of different types of point mutations.
8.2 Distinguish the detection of known mutations from scanning for unknown mutations.
8.3 Explain the use of immunoassays for detecting gene mutations and altered proteins.
8.4 Explain the concept of separation and detection by HPLC, chromatography, and spectrometry.
8.5 List and describe hybridization analyses to detect DNA mutations.
8.6 Describe enzymatic cleavage methods for DNA sequence analysis.
8.7 Determine which detection methods are appropriate for the screening of new mutations or the detection of previously identified mutations.
8.8 Describe gene mutation nomenclature for expressing sequence changes at the DNA, RNA, and protein levels.

Gene mutations include deletions, insertions, inversions, translocations, and other changes that can affect one base pair to hundreds or thousands of base pairs. Large differences in DNA sequence will likely have a significant effect on protein sequence. Alterations of a single or a few base pairs, or **point mutations,** will have a range of effects on protein sequence. A difference of one or a few base pairs may or may not change the encoded amino acid.

Point mutations are increasingly analyzed by sequencing methods. Sequencing not only directly detects the mutated base or bases but also provides the context of neighboring bases. Some sequencing methods also provide the percentage of variant alleles compared to a reference or normal sequence. Different sequencing technologies have limitations with respect to certain types of mutations, such as structural chromosomal abnormalities and copy-number variations. Next-generation sequencing (NGS) detection of germline variants in genetic testing also requires confirmation of the variant by Sanger sequencing or by other means. Phenotypic alterations in protein structure can only be predicted from the nucleotide sequence. Frequently occurring variants may be easily analyzed by simple and inexpensive single-mutation tests. Therefore, standard biochemical, cytogenetic, and molecular methods are still important.

TYPES OF GENE MUTATIONS

Because there is more than one codon for most of the amino acids, DNA sequence changes do not necessarily change the amino acid sequence. This is an important concept for interpreting the results of mutation analyses. Substitution of one nucleotide with a different nucleotide may be **silent,** that is, not change the amino acid sequence (Table 8.1). **Conservative substitutions** may change the amino acid sequence, but the replacement and the original amino acid have similar biochemical properties (e.g., leucine for valine), and the change will not drastically affect protein function. In contrast, a **nonconservative substitution** results in the replacement of an amino acid with a biochemically different amino acid (e.g., proline for glutamine), which changes the biochemical nature of the protein. A **nonsense substitution** mutation terminates proteins prematurely when a nucleotide substitution produces a **stop codon** instead of an amino acid codon. About 11% of disease-related gene lesions are nonsense mutations.[1]

Insertion or deletion of other than a multiple of three nucleotides results in a frameshift mutation, throwing the triplet code out of frame. The amino acids in the chain after the frameshift mutation are affected because the triplet code will include new combinations of three nucleotides. The genetic code is structured such that frameshifts often terminate protein synthesis prematurely because a stop codon appears sooner in the out-of-frame coding sequence than it would in an in-phase reading frame.

Nonconservative, nonsense, and frameshift mutations will generate different phenotypes, depending on where they occur along the protein sequence.[2] Point mutations in the 3′ end of a coding region may have minimal consequences, whereas mutations at the beginning of a coding sequence (5′ end or amino terminal end of a protein) are more likely to result in drastic alterations or even effective deletion of the protein-coding region.

TABLE 8.1 Types of Point Mutations

DNA Sequence	Amino Acid Sequence	Type of Mutation
ATG CAG GTG ACC TCA GTG	M Q V T S V	None
ATG CAG GTT ACC TCA GTG	M Q V T S V	Silent
ATG CAA GTG ACC TCA GTG	M Q L T S V	Conservative
ATG CCG GTG ACC TCA GTG	M P V T S V	Nonconservative
ATG CAG GTG ACC TGA GTG	M Q V T ter	Nonsense
ATG CAG GTG AAC CTC AGT G	M Q V NLS	Frameshift

These factors are important when interpreting the results of mutation analyses. Merely finding a change in a test DNA sequence does not guarantee an altered phenotype. Furthermore, point mutations detected by NGS methods over large sequence regions have to be screened to distinguish among silent, **conservative,** and nonconservative changes. For inherited or recurring diseases, the specific change in the DNA may be ascertained from a family history or previous analysis.

DETECTION OF GENE MUTATIONS

Some, mostly inherited, disease-associated sequence changes in DNA occur frequently at the same genetic location, for example, for the factor V Leiden mutation and the hemochromatosis C282Y, H63D, and S65C mutations. Also, increasing numbers of specific single-nucleotide polymorphisms (SNPs) are being mapped close to disease genes. These changes, although outside of the disease gene, are detected as specific sequence changes frequently inherited along with the disease phenotype.

Advanced Concepts

The nature of the genetic code is such that frameshift mutations lead to a termination codon within a small number of codons. This characteristic might have evolved to protect cells from making long nonsense proteins. It is a useful parameter to identify open reading frames in DNA.

Some diseases are associated with many possible mutations in a single gene. For instance, there are more than 600 disease-associated mutations in the cystic fibrosis transmembrane regulator *(CFTR)* gene, and more than 2,000 cancer susceptibility mutations have been reported in the *BRCA1* and *BRCA2* genes. Furthermore, unknown numbers of gene mutations are yet to be discovered. Detection of mutations in large genes requires screening across thousands of base pairs to detect a single altered nucleotide. Advances in sequencing technology allow genome-wide scanning for yet-unreported mutations. Whole-genome sequencing, however, is currently a complex approach, especially with regard to interpretation.

In some cases, biochemical-based methods are used to detect or quantify the altered protein product. These methods provide a more direct analysis of affected proteins and their structure and function.

Biochemical Methods

Biochemical methods are used to directly analyze the change in protein structure or function rather than to search for potential point mutations. This type of testing is also useful for metabolic defects where several genes are involved in the disease phenotype and the actual protein or amino acid alterations can be detected using biochemical methods. Commonly used biochemical methods include immunoassays that detect the presence of hormones, drugs, antibodies, cancer biomarkers, and other metabolites in blood, urine, or other biological fluids. Immunohistochemistry (IHC) is a longstanding method that allows detection of protein abnormalities in situ so that the tissue and intracellular location of mutant proteins (or lack thereof) can be observed. In addition, automated methods based on high-performance liquid or gas chromatography (GC) and mass spectrometry (MS) are also frequently used; these are described later in the chapter.

Enzyme Immunoassays

Immunoassays are flexible, potentially high-throughput methods for detecting a variety of analytes. These methods involve the use of specific antibodies or other ligands to detect the presence of the target molecules (Fig. 8.1). For this type of assay, plate wells, strips, or capillaries coated with capture antibody are exposed to test fluid (serum, plasma, or urine) that is diluted in the appropriate assay buffer. If the analyte is present in the test fluid, it will bind to the immobilized antibody. After rinsing away unbound material, detection antibodies covalently linked to alkaline phosphatase or horseradish peroxidase are introduced. The unbound antibody conjugates are washed away, and a substrate is added that will generate chemiluminescence, fluorescence, or color signals. Variations of this assay include immobilization of the antigen to detect target antibodies in the test fluid. The antibodies are then detected using anti-human antibody conjugates. This test is useful in the detection of antibodies against infectious agents.

Enzyme immunoassay (EIA) liquid handling techniques have been effectively automated. Turbidimetry (latex agglutination), chemiluminescence, and magnetic particle methods are supported on EIA analyzers. Most work with the 96-well plate format, and there are

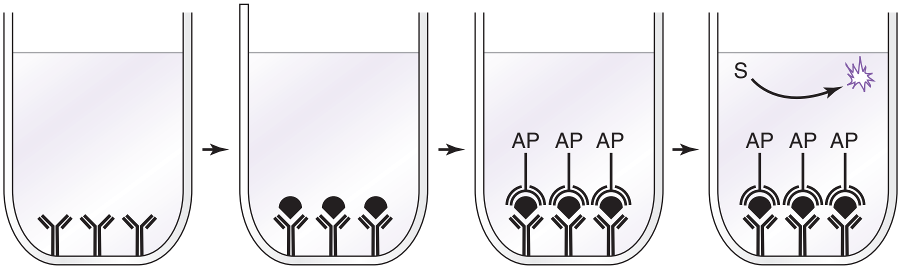


FIGURE 8.1 Enzyme immunoassay formats for direct antigen detection. Antibody specific for the target analyte is immobilized in plate wells. If present, antigen binds to the antibody and is detected with a secondary antibody conjugated to enzyme (AP, alkaline phosphatase).

preoptimized reagent sets designed for a wide variety of analytes.

Historical Highlights

EIAs and enzyme-linked immunosorbent assays (ELISAs) are modifications of an earlier technique called radioimmunoassay (RIA) that was first reported as a method to detect insulin in plasma.[3] A significant concern for RIA and its later variations was the radioactive hazards from iodine 131 (^{131}I), even when it was replaced with less toxic forms. In 1970, chemical methods were designed that coupled enzymes (alkaline phosphatase and others) with the antibodies to the analyte. These enzymes would then produce a signal from a color or light-producing substrate. In 1971, Engvall and Perlmann reported the use of ELISA to detect immunoglobulin G in rabbit serum.[4] That same year, van Weemen and Schurs demonstrated the use of EIA to quantify human chorionic gonadotropin concentrations in urine.

Immunohistochemistry

Historical Highlights

Paul Ehrlich coined the term “antibody” in 1891. In 1940 immunofluorescence staining on frozen sections based on antigen–antibody interactions was described by Coons.[5] Thirty-four years later, Taylor and Burns performed IHC on routinely processed fixed tissues. Originally performed using antisera containing polyclonal antibodies from a natural reaction against antigen, IHC was improved with the use of monoclonal antibodies. Monoclonal antibodies (mAbs) are produced by fusion of a single antibody-producing cell with an immortal cell to replicate and produce many copies of the same antibody. Köhler and Milstein developed this hybridoma technique in 1975.[6] Unlike polyclonal antibodies that target multiple epitopes with varying specificity, use of mAbs resulted in less nonspecific staining and better image quality.

Immunohistochemistry (IHC) is performed on thin (<5 micron) slices of fixed or 5- to 15-micron slices of frozen tissue. Fixation of tissue in formalin preserves tissue morphology. Sections from fixed tissue embedded in paraffin are made on a microtome. Fixation can also affect tissue antigens, however, altering or covering some **epitopes** (targets for mAb binding). Enzyme digestion with protein-digesting enzymes (proteinase K, trypsin, chymotrypsin, pepsin, pronase) or heating tissue sections in water or buffer can uncover antigen epitopes. This treatment is called **antigen retrieval.** To avoid the effects of formalin fixation, snap frozen tissue may be used. Tissue quickly frozen in isopentane (liquid at −160°C) can be sectioned in a **cryostat,** which allows cutting of frozen tissue into 4- to 15-micron sections inside of a chamber held at −20°C. Frozen sections can then be fixed in acetone, rather than formalin, to preserve structural morphology as well as the target epitopes. The sections are then dried and stored frozen. Dried sections can be rehydrated in phosphate-buffered saline, which also will remove the embedding material used to hold and position the tissue in the cryostat.

Substances such as endogenous peroxidase, fluorescence, or nonspecific antibodies in the tissue may interfere with IHC results. Blocking the background staining with hydrogen peroxide, ultraviolet (UV) light, or 1% serum will prevent background signals. Frozen sections can be treated with 0.1% sodium azide with 0.3% hydrogen peroxide. A blocking solution containing serum protein (albumin), detergent (Tween 20), and unlabeled antibodies will minimize nonspecific binding. Positive and negative controls are included with samples to ensure the adequacy and specificity of staining.

Imaging or microscopic observation of antibody binding requires a signal from the antibody. This signal can be fluorescent or colorimetric. To generate the signal, antibodies are covalently attached to fluorescent molecules or enzymes (horseradish peroxidase or calf intestine alkaline phosphatase). For fluorescent signals, the fluorescent molecules (fluorescein, Cy5, phycoerythrin) attached to the bound antibodies will emit signals when the fluor is excited. For colorimetric signals, a substrate solution is added to the bound antibodies. The substrate is oxidized by horseradish peroxidase or alkaline phosphatase, leading to a color reaction. Color generation will depend on the substrate; however, red or brown IHC staining is the most frequently used. Staining with more than one color may be achieved using sequential staining with different antibodies and substrates.

For direct antibody staining, the fluors or enzymes are attached to the antibody that directly binds the target molecules in the tissue (primary antibody). This method is faster than the alternative indirect staining, but it has limited signal intensity. In indirect staining, the primary antibody is not attached to the signaling molecule. A second or secondary antibody binds the primary antibody (usually from a different species, such as rabbit, mouse, or goat), carries the signal. Indirect staining may require additional blocking with unlabeled antibodies or IgG protein to avoid nonspecific binding to the secondary antibodies. The binding of a single primary antibody by multiple secondary antibodies amplifies the signal, allowing greater sensitivity of detection (Fig. 8.2).

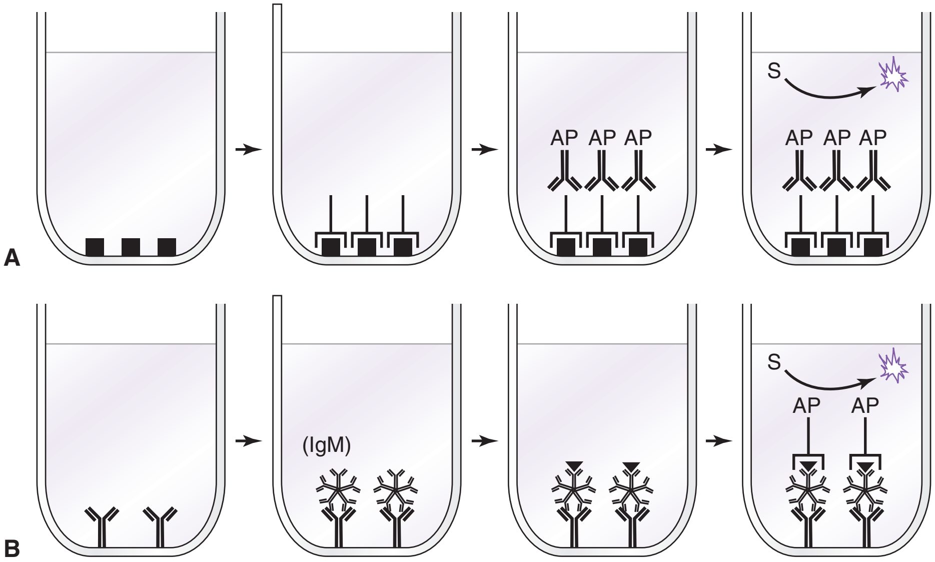


FIGURE 8.2 For antibody detection, antigen is immobilized and bound to specific antibody, if present, which is detected with AP-conjugated secondary antibody (A). Antigen may also be detected with immobilized antibodies to IgM or IgG, which will capture antigen for detection with enzyme conjugate (B).

Staining antibodies are diluted from 100 to several hundred-fold or more in blocking solution. The first blocking solution (without staining antibodies) is removed, and the staining solution is applied to the tissue for 10 to 30 minutes. Stained sections are rinsed with a buffer (e.g., Tris-buffered saline with 0.05% Tween 20) and visualized microscopically. For dual staining, the blocking, staining, and washing procedure is repeated with a second antibody. Visualization of fluorescent signals is direct, but for color developed through oxidation of a substrate, the substrate solution is placed on the section for 10 to 60 minutes. After washing, the slides can be counterstained to visualize unstained structures. The stained sections are then dehydrated with 100% ethanol and air-dried.

IHC has been a standard in pathology for many years.[7] It provides the advantage of integrating target detection, localization, and quantification in the context of tissue morphology. Targeted therapeutic agents have increased the use of IHC to assess tissue expression of the target molecules. This guides treatment strategy at a relatively low cost compared with other methods.

High-Performance Liquid Chromatography (HPLC)

Liquid chromatography was first designed to separate pigments extracted from plants using a solvent that was poured into a glass column packed with chalk and alumina. The Russian botanist Mikhail S. Tswett named this process chromatography from the Greek words *chroma,* meaning "color," and *graph,* meaning "writing."[8]

HPLC is the basis for separation/analysis instruments such as amino acid analyzers. Migration rates of molecules differ with varying combinations of HPLC components and detectors. HPLC consists of two phases, a mobile phase or solvent that flows through a stationary phase or solid support. The solvent chemistry is selected so that it interacts more strongly with target molecules to slow their migrations or more weakly to speed their migrations. Organic solvent in the mobile phase may be the same throughout the column **(isocratic)** or of increasing strength (gradient). The stationary phase will also interact with the sample and affect migration speed. Molecules can thus be identified from normal migration patterns.

Stationary phases of HPLC differ depending on the application. Size-exclusion columns are comprised of porous beads that exclude larger molecules and retain smaller molecules inside of the beads so that the larger molecules are eluted faster than the smaller molecules (Fig. 8.3). Normal and reverse-phase columns separate on the basis of hydrophilicity, with lipophilic molecules eluting faster, or hydrophobicity, with lipophilic molecules eluting slower. In ion exchange, ions in the sample are retained as counter ions to charged groups that are permanently attached to the stationary phase. Changing the chemical properties of the mobile phase will selectively elute the trapped sample ions. Another selective solid phase, affinity, is designed to immobilize specific ions while other ions are washed through. The selected ion is eluted by changing the mobile-phase conditions.

When a test sample is dissolved in the mobile phase, it is introduced to HPLC by injection with a syringe through an injection port in the column. Pumps will force the sample through the column, and elutions are monitored by a detector that produces a readout of signal peaks as the sample components elute. There are many types of detectors, depending on the characteristics of the sample, including light scattering, fluorescence, refractive index, UV light absorption, and MS. The retention time and size of peaks indicate the type and amount of sample components. HPLC may be used to separate nucleic acids as well as proteins.[9]

Higher powered ultra-HPLC (UHPLC) columns have been devised to increase resolution and lower separation time while using less solvent.[10] UHPLC uses smaller-diameter columns and smaller stationary-phase particles than HPLC, with faster flow rates, up to 5 mL/min. UHPLC is not recommended for complex, unfiltered samples.

Gas Chromatography

In GC, which is an automated method of analysis, the mobile phase is an inert gas, and the stationary phase is a high-boiling-point liquid that is absorbed to an inert solid support in the column (Fig. 8.4). The sample is introduced to the column and vaporized into a gas. The inert gas carries the sample through the column. The strength of the interaction or dissolution of the sample components into the liquid phase will result in varying

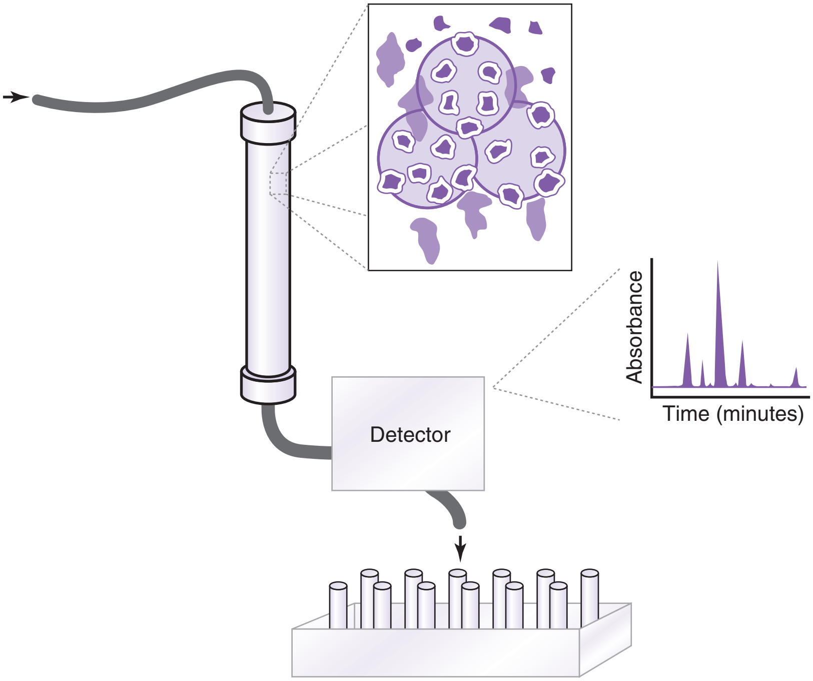


FIGURE 8.3 Liquid chromatography is the separation of molecules in solution through interaction with a solid support in the column. Bead or particulate matrices have small spaces that hold and impede movement of small particles. Larger particles migrate around the beads. Separated molecules are detected directly or collected fractions are further analyzed by MS.

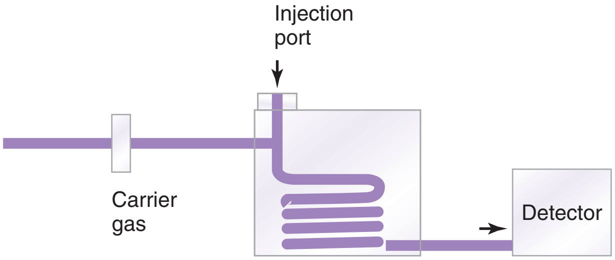


FIGURE 8.4 Gas chromatography (gas-liquid chromatography) is the separation of vaporized sample through a column of inert carrier gas and liquid stationary phase that differentially adsorbs molecules. The sample is injected through the injection port, and the separated molecules are recorded by the detector.

retention times in the column. The effluent from the column is monitored by a detector, such as a flame ionization detector, that is sensitive to organic molecules. Alternatively, GC may be coupled to a mass spectrometer. GC is used for detection of drugs and poisons and their metabolites in biological samples.[11,12] Coupled with MS, GC is being used to detect biomarkers of disease.[13,14]

Mass Spectrometry

A mass spectrometer converts molecules to ions that can be moved in a magnetic field based on their charge and mass. In this automated method, an ion source sends high-energy electrons that hit the target sample molecules, separating them into ions, usually cations with the loss of one or two electrons. The ions themselves may further fragment into smaller ions and neutral particles. This collection of particles is accelerated and focused into a beam through a magnetic field that deflects the ions according to their mass and charge. By varying the strength of the magnetic field, the ions are aimed at a detector. The readout of the instrument is a spectrum with the mass/charge value on the *x*-axis and the abundance of the ion on the *y*-axis (Fig. 8.5). Molecules can therefore be identified by their characteristic spectrum or set of peaks.

Two ionization methods have been developed for MS of large biomolecules such as proteins: electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI). Both result in ion formation with limited loss of sample integrity. In ESI, a test sample is converted into a fine spray of charged droplets that are

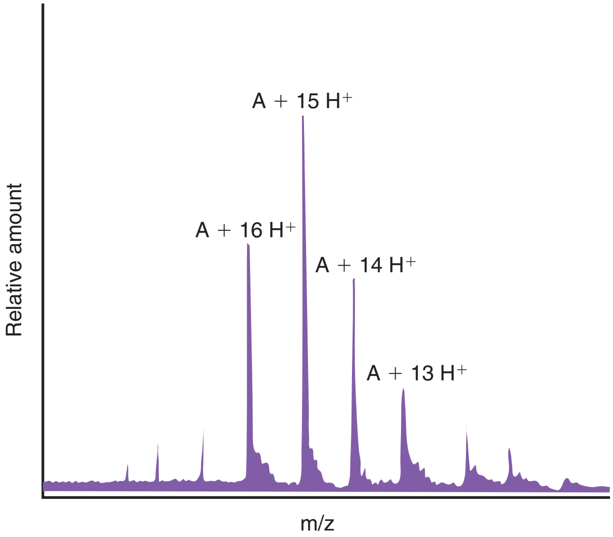


FIGURE 8.5 In ESI spectrophotometry, multiple ionized species from one protein are separated by mass and charge. The results are plotted as mass/charge ratio (m/z) and relative abundance.

electrostatically directed to the mass spectrometer inlet through a vacuum where ions are released and drawn through electrostatic lenses to the MS inlet. MALDI[15] produces ions by firing a laser pulse into the sample coated with a matrix (organic compound). Desorption produces a positive ion from the sample molecule (Fig. 8.6). The ion flies to the detector at a speed (time of flight) based on its mass and charge. High-molecular-weight molecules are identified using time of flight with MALDI ionization (MALDI-TOF). If the ions are allowed to drift in the flight tube, they will separate according to mass/charge ratio so that the lighter ions travel faster and reach the detector before the heavier ions do (Fig. 8.7). An extension of MALDI, surface-enhanced laser desorption/ionization (SELDI), combined with time of flight (SELDI-TOF) offers flexibility in the identification and quantification of peptides.

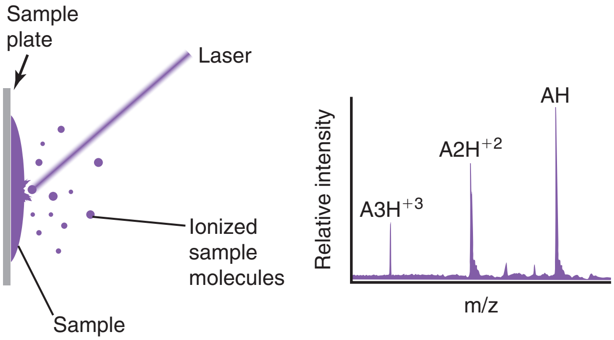


FIGURE 8.6 In MALDI spectrophotometry, sample is adhered to the matrix on a sample plate, ionized (protonated), and released by laser bombardment. The ionized molecules are separated in an electric field by mass and charge. A sample molecule (A) will have multiple protonated species (AH, A2H, A3H).

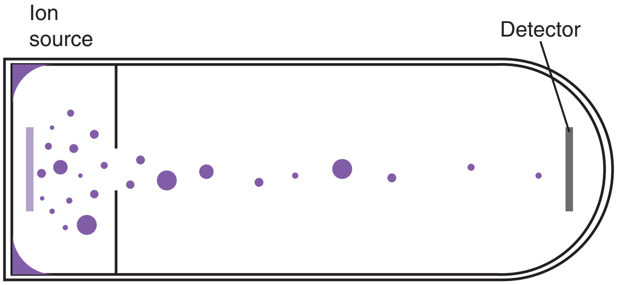


FIGURE 8.7 In MALDI-TOF spectrophotometry, the ionized molecules are accelerated at a fixed point and allowed to drift through the flight tube to the detector. The time of flight is determined by the mass/charge ratio of the molecules. DNA sequence variants are detected as extension products of different sizes (mass).

Advanced Concepts

It is possible to analyze almost any class of molecule with mass spectrometry techniques. Further information about the basic structure of a compound is gained by three-dimensional quadrupolar electric fields, tandem MS (MS/MS), and hybrid instruments such as MALDI-TOF/TOF. The resolution of these instruments is sufficient to identify amino acid sequences in peptides.

Nucleic acids can also be analyzed by MS. Single-base-pair changes can be detected using a primer extension technique. In this method, primers are bound to test DNA template just adjacent (5′) to the base position to be analyzed. Extension of the primer with a dideoxynucleotide (ddNTP) terminates the extension at one base and changes the mass and charge of the primer. The primer extension will only occur if the incoming ddNTP is complementary to the template base next to where the primer is bound. By MALDI-TOF, the mass of the extended primer indicates the sequence or variant at the template site. Software automatically translates the mass into a genotype. Multiple variants can be analyzed simultaneously in a 384-well plate format.

Nucleic Acid Analyses

Biochemical detection and/or characterization of proteins using the biochemical methods just described may be part of a clinical chemistry laboratory menu or performed as molecular diagnostics. Mutation detection by analysis of nucleic acids is considered the classical molecular methodology.

Nucleic acid analysis is performed on a variety of specimen types. Inherited mutations are detected from the most convenient and noninvasive specimen material, such as blood or buccal cells. Somatic mutations are often more challenging to find because cells harboring mutations may be only a small fraction of the total specimen that consists of mostly normal cells. Polymerase chain reaction (PCR) amplification, which is part of many procedures, has simplified mutation detection, especially from limiting specimens. PCR or other amplification methods, however, must be performed under conditions that minimize the introduction of mutations in the course of amplification. Although advances in genomic analysis have lowered the cost and increased the resolution and throughput of mutation detection, methods targeting single variants remain a rapid and cost-effective approach.

Interpretation of the results of such targeted mutation analyses may also be challenging. Mutation scanning by methods that do not indicate the primary sequence change do not differentiate between silent, conservative, and nonconservative mutations. The actual effect on phenotype is left to posttest interpretation of supporting clinical data and patient family history. Mutations discovered through this type of scanning may be subjected to sequence analysis to confirm and further characterize the mutated region.

Although it is the most definitive method for detecting mutations, DNA sequencing may not be appropriate, especially for high-throughput procedures. A number of techniques have been designed for the detection of DNA mutations, from single-base-pair changes to large chromosomal rearrangements, without having to determine the primary DNA sequence. Some of these methods are described in the following sections.

Sequence detection methods can be generally classified according to three broad approaches: hybridization-based methods, sequence (polymerization)-based methods, and enzymatic or chemical cleavage methods. Brief descriptions of representative methods are presented in the following sections. The methods selected are currently used or proposed for use in clinical applications. A summary of the methods discussed in this chapter is shown in Table 8.2.

Hybridization-Based Methods

Single-Strand Conformation Polymorphism (SSCP)

SSCP is based on the preference of DNA (as well as RNA) to exist in a double-stranded, rather than single-stranded, state.[16,17] In the absence of a complementary strand, nucleic acids form intrastrand duplexes to attain as much of a double-stranded condition as possible. Each folded strand forms a three-dimensional structure, or **conformer,** the shape of which is determined by the primary sequence of the folded strand. The migration of the single-stranded conformers in polyacrylamide gels under precisely controlled denaturing and temperature conditions distinguishes sequence variants.

For SSCP, dilute concentrations of short, double-stranded PCR products, optimally 100 to 400 base pairs (bp) long, are denatured (e.g., in 10 to 20 mM NaOH, 80% formamide for 5 minutes at 95°C; or 10 to 20 mM NaOH, 0.004 mM EDTA, 10% formamide for 5 minutes at 55°C to 60°C), followed by rapid cooling. Because the diluted single strands cannot easily find their homologous partners, they fold by intrastrand hybridization, forming three-dimensional conformers. The shape of the conformer depends on the complementary nucleotides available for hydrogen bonding and folding. A single-base difference in the DNA sequence can cause the conformer to fold differently. These conformers are resolved in a polyacrylamide gel or by capillary electrophoresis with strict temperature control. The speed of migration depends on the shape as well as the size of the conformer. Differences in the shape of the conformers (kinks, loops,

TABLE 8.2 Summary of Mutation Detection Methodologies

Method*	Target[†] (bp)	Accuracy[‡] (%)	Specificity[§] (%)	Sensitivity[‖] (%)
Sequencing	>1000	100	100	10–20
SSCP	50–400	70–100	80–100	5–20
ASO	Defined	100	90–100	5–20
HR-MCA	Defined	95–100	95–100	1–5
DHPLC	50–1000	95–100	85–100	5–20
Array technology	Defined, multiplex	95–100	80–100	1–5
SSP	Defined	98–100	95–100	0.0005
Allelic discrimination	Defined	95–100	90–100	0.0001
PCR-RFLP	Defined	100	100	0.01–1
Cleavase	Defined, multiplex	100	95–100	

*See text. Data are from methods done under optimal conditions.

[†] Optimal length of sequence that can be screened accurately; defined methods target a single nucleotide or site; multiplex methods target multiple defined types in the same reaction.

[‡] Concordance with direct sequencing or other assays reported in the references.

[§] True positive detection of mutations without concurrent false positive.

[‖] Detection of one mutant target in a background of normal targets.

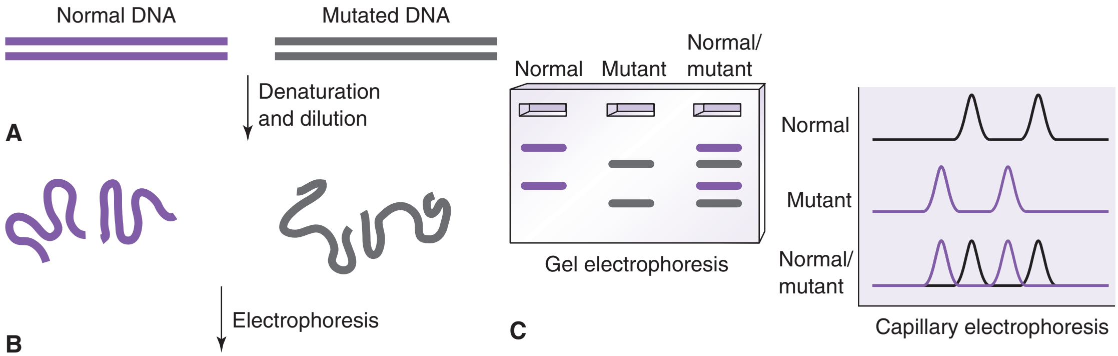


FIGURE 8.8 Single-strand conformation polymorphism analysis. Double-stranded PCR products (A) of normal or mutant sequences are denatured and form conformers (B) through intrastrand hydrogen bonding. These conformers can be resolved (C) by gel *(left)* or capillary *(right)* electrophoresis.

bubbles, and tails) are caused by sequence differences in the DNA single-strand nucleotide sequence (Fig. 8.8). The band or peak patterns are detected by silver stain, radioactivity, or fluorescence. The use of low concentrations avoids the renaturation of homologous partners. As a consequence, less sensitive stains such as ethidium bromide are not often used for this assay. Band or peak patterns different from those of normal sequence-control conformers prepared simultaneously with the test conformers indicate the presence of gene mutations.

SSCP is reported to detect 35% to 100% of putative mutations.[18] The assay can be sensitive enough to detect mutations in samples containing as low as 5% potentially mutant cells,[19] although specimens that are at least 30% potentially mutant cells produce more reliable results. This requirement is satisfied in inherited mutations because at least 50% of cells of a specimen will potentially carry a mutation. However, for somatic mutations, such as the analysis of tumor cells, the potentially mutant cells may be mixed with or surrounded by a vast majority of normal cells or tissue. Therefore, a cell suspension that is at least 30% tumor cells or a microdissection of solid tumor tissue from fixed or frozen sections is recommended.

Advanced Concepts

Because SSCP worked more accurately on some genes than others, modifications of the SSCP procedure were developed, for instance, using RNA instead of DNA **(**RNA-SSCP or rSSCP**)**[20,21] or using restriction endonuclease fingerprinting **(REF-SSCP).**[22] The rSSCP and REF-SSCP methods, although more sensitive, were more difficult to interpret and were not put into general use.

Allele-Specific Oligomer Hybridization

Allele-specific hybridization, or allele-specific oligomer hybridization (ASO), utilizes the differences in the melting temperatures of short sequences of about 20 bases with one or two mismatches and those with no mismatches. Synthetic single-stranded probes with the normal or mutant target DNA sequence are used for this assay. At specific annealing temperatures and conditions (stringency), a probe will not bind to a near complementary target sequence with one or two mismatched bases, whereas a probe with a perfect complementary sequence will bind. ASO uses an immobilized target and a labeled probe in solution. This dot blot method was used to test for known, frequently occurring mutations, for example, in the *BRCA1* and *BRCA2* gene mutations frequently observed in inherited breast cancer[23] and the *p16* gene mutations in familial melanoma.[24] ASO analysis also may be carried out as a reverse dot blot in a 96-well plate format, similar to the capture probe methods developed for detection of *Chlamydia trachomatis* and *Mycobacterium tuberculosis*. For mutation analysis, mutant or normal probes were immobilized on a membrane. The sequence to be tested was amplified by PCR with one regular and one biotinylated primer. The denatured biotinylated products were then exposed to the immobilized probes under conditions set so that only the exact complementary sequences would hybridize. Bound probes were detected with a conjugated horseradish peroxidase-antibiotin Fab fragment and exposure to a chromogenic or chemiluminescent substrate. Generation of a color or light signal indicated the binding of the test DNA to the normal or mutant probe. ASO has been applied to human leukocyte antigen (HLA) typing, but it has now been replaced with microarray and bead array methods and, more recently, by next-generation sequencing.[25]

Melt-Curve Analysis

Like ASO, **melt-curve analysis (MCA)** exploits the sequence- and stacking-directed denaturation characteristics of DNA duplexes.[26] The method is a postamplification step of real-time PCR. PCR amplicons generated in the presence of a DNA-specific fluorescent dye, such as ethidium bromide, SYBR green, or LC green, are heated at a rate of about 0.3°C/sec. The dyes, specific for double-stranded DNA, initially yield a high signal because the DNA is mostly double stranded at the low temperature (50°C to 60°C). As the temperature rises, the DNA duplexes begin to separate into single strands, losing dye accordingly. The fluorescent signal gives a pattern as shown in Figure 8.9. Sequence differences result in

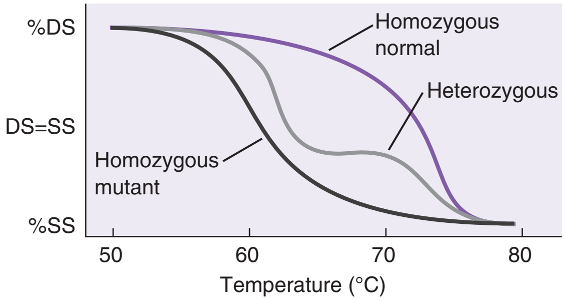


FIGURE 8.9 MCA of homozygous mutant, heterozygous, and normal PCR products. As the temperature increases, double-stranded (DS) DNA denatures into the single-stranded (SS) state.

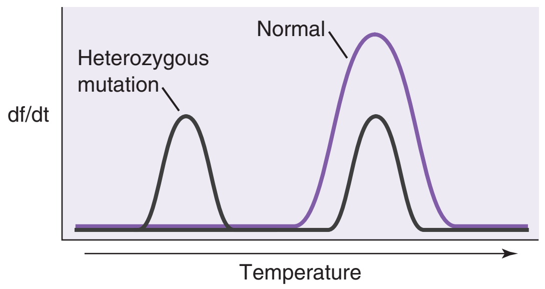
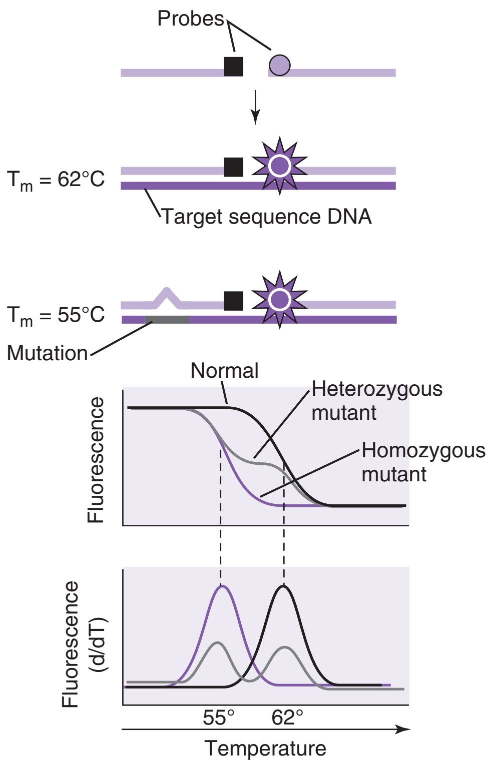


FIGURE 8.10 A plot of the derivative of the fluorescence data (*df/dt*) versus temperature shows the inflection point of the melt curve as a peak at the T_m of the test sequence. A normal homozygous sample should have a T_m that can be distinguished from that of the mutant sequence.

different melting characteristics and melting temperatures (T_ms; where there are equal amounts of double- and single-stranded DNA) for each sequence. The T_m is illustrated as a peak, plotting the derivative (speed of decrease) of fluorescence versus temperature. The results are interpreted by the temperature peak placement with respect to the temperature on the *x*-axis. Specimens with identical sequences should yield overlaying peaks at the expected T_m, whereas specimens containing different sequences will yield two or more peaks at different temperatures (Fig. 8.10).

The nonspecific dyes used in MCA are not sequence specific and therefore do not distinguish between the target amplicon and extraneous products that may occur in the PCR reaction, such as primer dimers or misprimed amplicons. Although the target sample should be identifiable by its T_m, such unintended products can complicate the melt curve and confuse interpretation.

The specificity of MCA is increased by using high-resolution melt-curve analysis (HR-MCA),[27,28] which uses fluorescent resonance energy transfer (FRET) probes that hybridize next to one another across the sequence position being analyzed. The probes fluoresce only when bound to the target sequence because FRET fluorescence relies on the transfer of energy from a donor fluorescent molecule (fluor) on one probe to an acceptor fluor on the other probe. As the temperature increases, the probes dissociate at a specific T_m. When the probes dissociate from the target, the donor is no longer close to the acceptor, and the fluorescence drops (Fig. 8.11). If the target sequence has a mismatch between the target and the probe, hydrogen bonding is perturbed between the two strands of the double helix. The mismatch decreases the dissociation temperature compared with matched or complementary sequences. A T_m lower than that of the probe and its perfect complement, therefore, indicates the presence of a mutation, or sequence difference, between the probe reference sequence and the test sequence.



FRET probes

FIGURE 8.11 MCA with FRET probes. A mismatch between the target and probe will lower the T_m of the duplex.

Advanced Concepts

PCR products smaller than 300 bp in size are preferred for MCA. The ability of the assay to distinguish sequence differences decreases with the increasing size of the PCR product.[27]

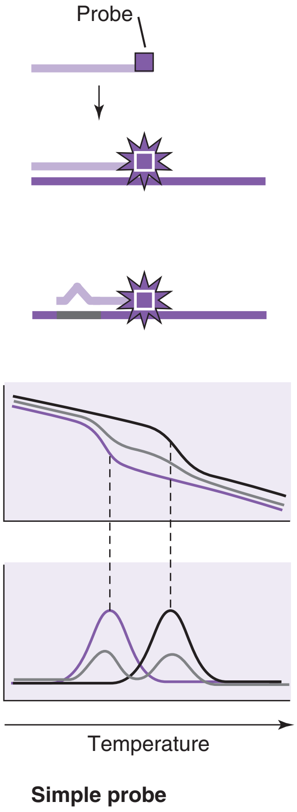


FIGURE 8.12 MCA with a single probe. The binding of the single probe and fluorescence will decrease with increasing temperature. The rate of decrease will depend on the hybridized sequences.

FRET is most frequently performed with two probes; however, single-probe systems have been developed. The single probe is designed to fluoresce much more brightly when hybridized to the target. The fluorescence is lost on dissociation (Fig. 8.12). Another modification that is reported to improve the sensitivity of MCA is the covalent attachment of a minor groove binder (MGB) group to the probe. The MGB, dihydrocyclopyrroloindole tripeptide, folds into the minor groove of the duplex formed by hybridization of the terminal 5 to 6 bp of the probe with the template. This raises the melting temperature of the probe, especially one with high A/T content. The T_m of a 12- to 18-bp MGB conjugated probe was measured to have hybridization properties equivalent to that of a 25- to 27-bp non-MGB probe.[29]

Special instrumentation is required for MCA and HR-MCA. Thermal cyclers with fluorescent detection have melt-curve options that can be added to the thermal cycling program. Instruments that perform MCA only can process more samples per unit time than the thermal cycler systems. Melt-curve methodology has been proposed for a variety of clinical laboratory applications, such as detection of DNA polymorphisms and typing of microorganisms. Naturally occurring sequence variation must be considered in performing MCA because it can complicate the interpretation of the test.

Heteroduplex Analysis

Solution hybridization and electrophoresis of test nucleic acid fragments mixed with reference nucleic acid fragments can reveal mutations. To form heteroduplexes, nonidentical double-stranded DNA duplexes are heated to a temperature that results in complete denaturation of the double-stranded DNA (e.g., 95°C) and then slowly cooled (e.g., −1°C/4 to 20 sec). Heteroduplexes are formed when single strands that are not completely complementary hybridize to one another (Fig. 8.13). (Heteroduplexes are also formed when test PCR products amplified from genetically heterozygous specimens are denatured and renatured.) The heteroduplexes migrate differently than do **homoduplexes** through polyacrylamide or agarose gels. The presence of bands different from a homozygous reference control is indicative of mutations. Gel-based heteroduplex methods have been designed for HIV typing and hematological testing.

Heteroduplexes are also resolved by denaturing high-performance liquid chromatography (DHPLC). This version of **heteroduplex analysis** is performed on PCR products, ideally 150 to 450 bp in length. HPLC separation is performed on a 25% to 65% gradient of acetonitrile in triethylammonium acetate at the melting temperature of the PCR product. The heteroduplexes elute ahead of the homoduplexes as the denaturing conditions intensify. The migrating homoduplexes and heteroduplexes are detected by absorbance at 260 nm or by fluorescence. HPLC methods were reported to be more sensitive than gel methods, with greater capacity for screening large numbers of samples. Although HPLC analysis of heteroduplexes was evaluated as a mutation screening method, gel-based heteroduplex analyses are still routinely used in the clinical laboratory.

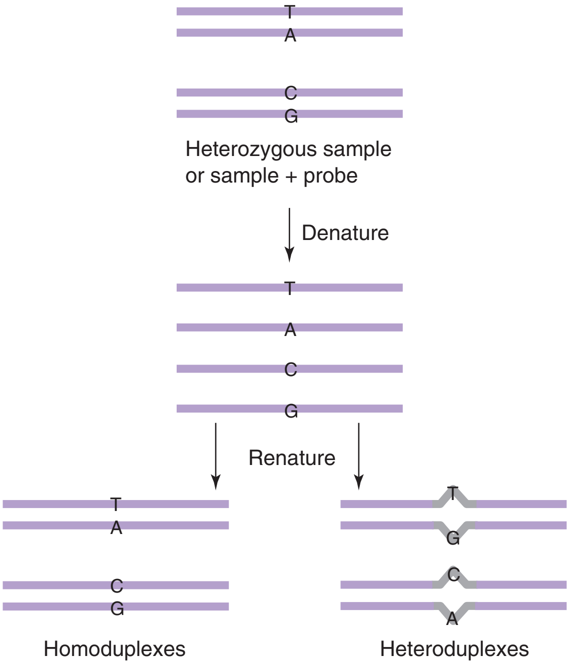


FIGURE 8.13 Heteroduplex analysis is performed by mixing sample amplicons with a reference amplicon, denaturing, and slowly renaturing. If the sample contains mutant sequences, a fraction of the renatured products will be heteroduplexes. These structures can be resolved from homoduplexes by electrophoresis.

Array Technology

Single-base-pair resolution by hybridization differences is achievable with high-density oligonucleotide arrays. These methods are similar to comparative genome hybridization but focus on single genes with higher resolution, as in ASO procedures. The advantage of array methods is the large number of inquiries (potential sequence mutations or SNPs) that can be tested simultaneously. For analysis, the test DNA is fragmented by treatment with DNase before binding to the complementary probes on the array. If the sample fragments are too large (not treated with DNase), a single-base-pair mismatch has minimal effect on hybridization because the fragment binds to multiple probes and the specificity of detection is lost. An example of one type of hybridization format, **standard tiling,** is shown in Figure 8.14. In this format, the base substitution in the immobilized probe is always in the 12th position from its 3′ end. Commonly occurring mutations are targeted in another type of format, **redundant tiling,** in which the same mutation is placed at different positions in the probe (at the 5′ end, in the middle, or at the 3′ end). After hybridization of the fluorescently labeled sample DNA, the fluorescent signal is read on a scanner with appropriate software, and the mutations are identified as indicated by which probes are bound. Although not performed routinely in clinical laboratories, a number of applied methods were developed

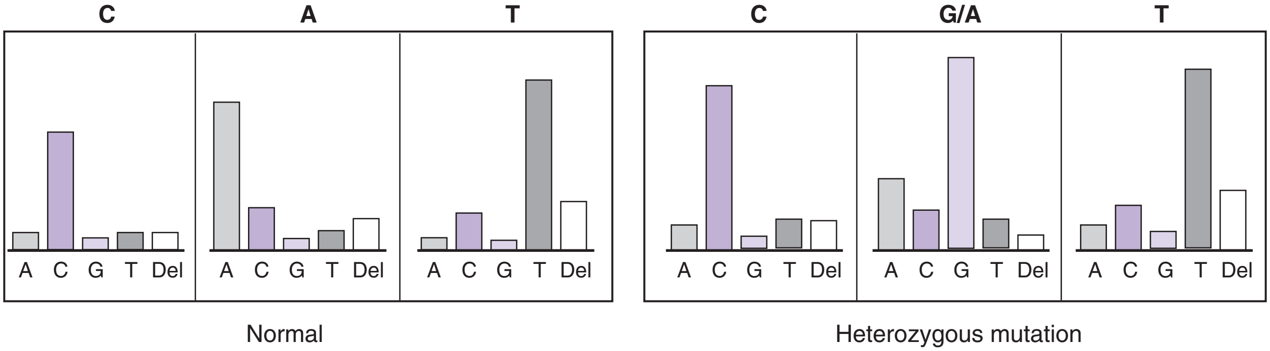


FIGURE 8.14 Mutation analysis of the *p53* gene by high-density oligonucleotide array analysis. Each sequence position is represented by 10 spots on the array, 5 sense and 5 antisense probes. The sequence binds only to its perfect complement probe. The illustration shows three adjacent sequence positions, CAT. Binding of the sample fragment is detected by increased fluorescence. A fragment with the normal sequence is on the right; a heterozygous mutation is on the left.

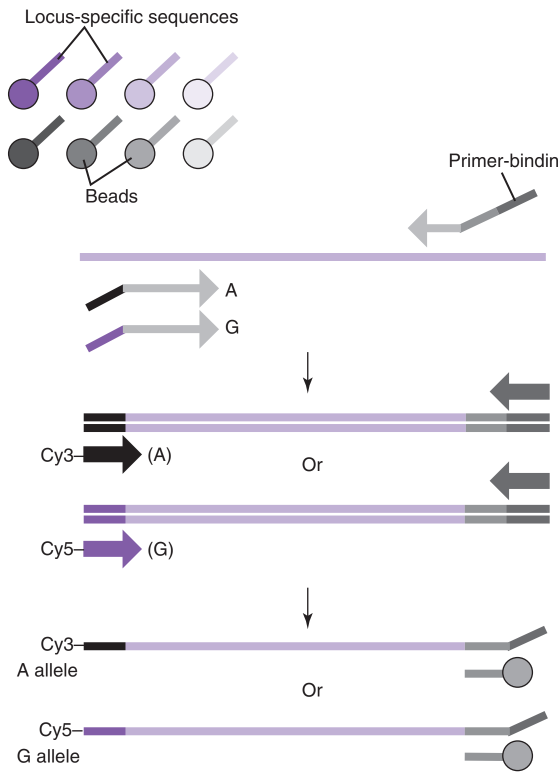


FIGURE 8.15 Bead array technology. Beads colored with distinct fluorescent dyes *(upper left)* are covalently attached to the probe sequences, with each color of bead attached to a probe representing a specific locus. In a sequence-specific PCR, test DNA is amplified with tailed primers. The tailed PCR products are amplified in a second reaction to generate labeled amplicons that will bind to specific beads, according to the gene locus. The combination of bead label and the hybridized amplicon label reveals whether there is a mutant or normal allele at that locus.

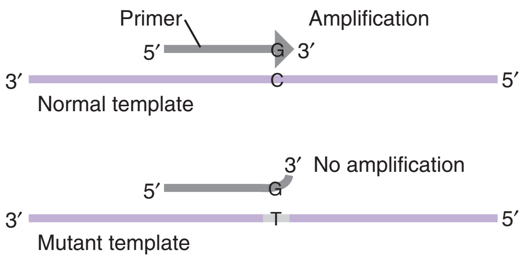


FIGURE 8.16 Sequence-specific primer amplification. Successful amplification will occur only if the 3′ end of the primer matches the template.

using high-density oligonucleotide and microelectronic arrays.[30,31]

Bead array technology utilizes sets of color-coded polystyrene beads in suspension as the solid matrix (Fig. 8.15). In an extension of a flow cytometry system,[32] 100 sets of beads are dyed with distinct fluorochrome mixes. Each set is coated with oligonucleotide probes corresponding to a genetic locus or gene region. In this technology, 10^5 or more probes are attached to each 3- to 6-micron-diameter bead. When labeled test samples are hybridized to the beads through complementary probe sequences, the combination of bead color and test label reveals the presence or absence of a mutation or polymorphism. The advantage of this arrangement is that multiple loci can be tested simultaneously from small samples. Up to 100 analytes are tested in a single well of a microtiter plate. This method requires a flow cytometry instrument that excites and reads the emitted fluorescence as the beads flow past a detector. This technology has been applied to antibody detection and infectious diseases and is used in tissue typing and in other clinical applications.

Sequencing (Polymerization)-Based Methods

Sequence-Specific (Primer) PCR

Sequence-specific (primer) PCR (SSP-PCR) is commonly used to detect point mutations and other SNPs. There are numerous modifications to the method, which involve the careful design of primers such that the primer 3′ end falls on the nucleotide to be analyzed. Unlike the 5′ end, the 3′ end of a primer must match the template perfectly to be extended by *Taq* polymerase (Fig. 8.16). By designing primers to end on a mutation, the presence or absence of product is interpreted as the presence or absence of the mutation.

In a modification of SSP-PCR, normal and mutant sequences are distinguished from one another by increasing the length of the normal or the mutant primer, resulting in differently sized products, depending on the sequence of the template (Fig. 8.17). Alternatively, primers can be multiplexed (Fig. 8.18). Multiplexed SSP-PCR was originally called amplification refractory mutation system PCR or tetra-primer PCR.[33] Sequence-specific PCR is routinely used for

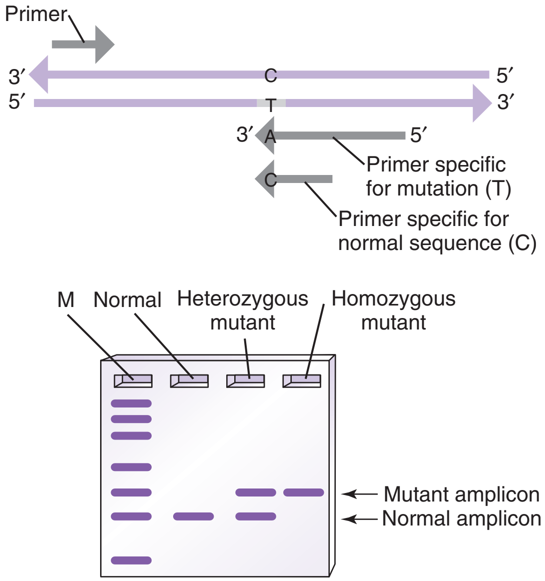

Primer
3′
5′
C
T
5′
3′
3′
A
5′
C
Primer specific for mutation (T)
Primer specific for normal sequence (C)
M
Normal
Heterozygous mutant
Homozygous mutant
Mutant amplicon
Normal amplicon

FIGURE 8.17 Allele-specific primer amplification of a C→T mutation. A longer primer is designed with the mutated nucleotide (A) at the 3′ end. This primer is longer and gives a larger amplicon than the primer binding to the normal sequence *(top)*. The resulting products can be distinguished by their size on an agarose gel *(bottom)*. First lane: molecular-weight marker; second lane: a normal sample; third lane: a heterozygous mutant sample; fourth lane: a homozygous mutant.

high-resolution HLA typing and for detection of commonly occurring mutations.

A high-throughput application of bead array technology[34,35] uses sequence-specific PCR (Fig. 8.15). In this assay, primers tailed at the 5′ end with locus-specific sequences and allele-specific sequences (A or G in Fig. 8.15) are used to amplify the test DNA. Each resulting PCR product will have an allele-specific sequence at one end and a locus-specific sequence at the other end. The PCR products are subsequently amplified in a second round using Cy3 or Cy5 (fluorescently) labeled 5′ primers, complementary to the sequences introduced by the tailed primers in the first round. These amplicons are then hybridized to the locus-specific sequences covalently attached to the bead array. The bead color (locus) combined with Cy3 or Cy5 fluorescence (allele) types the allele at each locus. This system is one of the technologies that was used in the Human Haplotype Mapping Project.

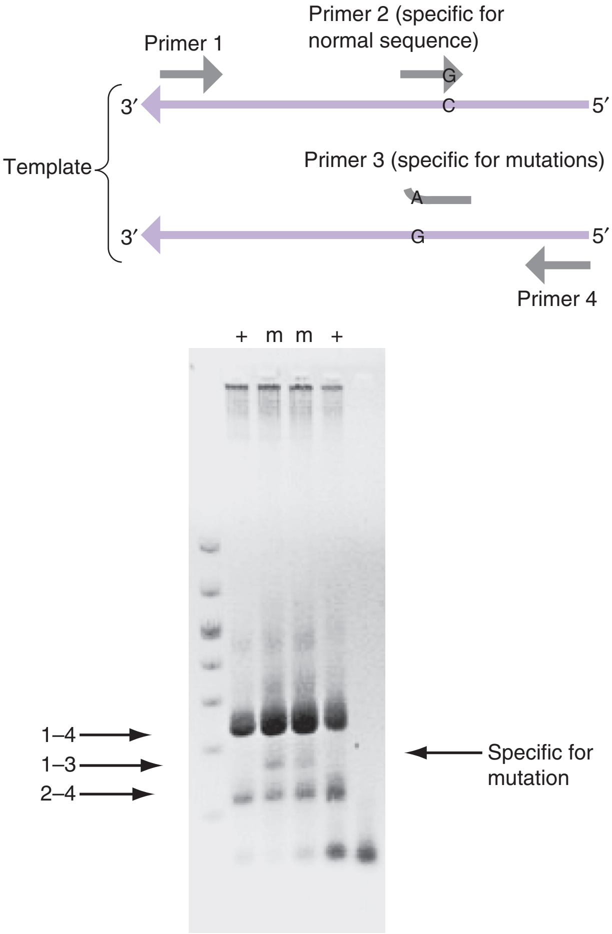

Primer 1
Primer 2 (specific for normal sequence)
G
3′
C
5′
Template
Primer 3 (specific for mutations)
A
3′
G
5′
Primer 4
+
m
m
+
1–4
1–3
2–4
Specific for mutation

FIGURE 8.18 Multiplex allele-specific PCR. The mutation (C→A) is detected by an allele-specific primer (3) that ends at the mutation. Primers 3 and 4 would then produce a mid-sized fragment (1–3). If there is no mutation, a normal primer (2) binds and produces a smaller fragment (2–4). Primers 1 and 4 always amplify the entire region (1–4).

Allelic Discrimination With Fluorogenic Probes

Thermal cyclers with fluorescent detection support allelic discrimination with fluorogenic probes. This method is a real-time PCR assay, using two probes labeled 3′ quencher molecules and different fluors on the 5′ ends (Fig. 8.18). Each probe is complementary to either the normal or mutant sequence. The hybridized probe is digested by the polymerase enzyme, releasing the reporter dye. The presence of the corresponding fluorescent signal indicates whether the test sequence is

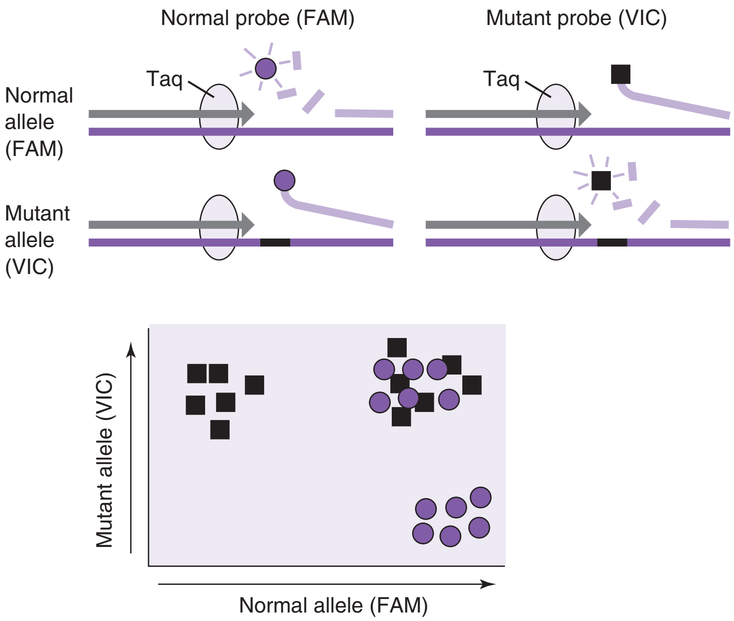


FIGURE 8.19 Allelic discrimination using fluorogenic probes. Probes, complementary to either the normal sequence *(left)* or the mutant sequence *(right),* are labeled with different fluors, for example, FAM and VIC, respectively. The *Taq* exonuclease functions only if the probe is matched to the sequence being tested. High FAM indicates a normal sequence, and high VIC indicates a mutant sequence. If both fluors are detected, the test sample is heterozygous.

normal or mutant, that is, whether the probe matched and hybridized to the test sequence. In the example shown in Figure 8.19, the probe complementary to the normal sequence is labeled with FAM dye; the probe complementary to the mutant sequence is labeled with VIC dye. If the test sequence is normal, FAM fluorescence will be high, and VIC fluorescence will be low. If the test sequence is mutant, VIC will be high, and FAM will be low. If the sequence is heterozygous, both VIC and FAM will be high. Negative controls show no VIC or no FAM. This assay has the advantage of interrogating multiple samples simultaneously and was proposed as a practical high-throughput laboratory method.[36,37]

Enzymatic and Chemical Cleavage Methods

Restriction Fragment Length Polymorphisms

If a mutation changes the structure of a restriction enzyme target site or changes the size of a fragment generated by a restriction enzyme, restriction fragment length polymorphism (RFLP) analysis can be used to detect the sequence alteration. To perform PCR-RFLP, the region surrounding the mutation is amplified, and the mutation is detected by cutting the amplicon with the appropriate restriction enzyme (Fig. 8.20). Mutations may inactivate a naturally occurring restriction site or generate a new restriction site so that digestion of the PCR product results in cutting of the mutant amplicon, but not a normal control amplicon or vice versa. Although straightforward, PCR-RFLP requires careful design, because rare polymorphisms have been reported to confound RFLP results. Several PCR-RFLP methods

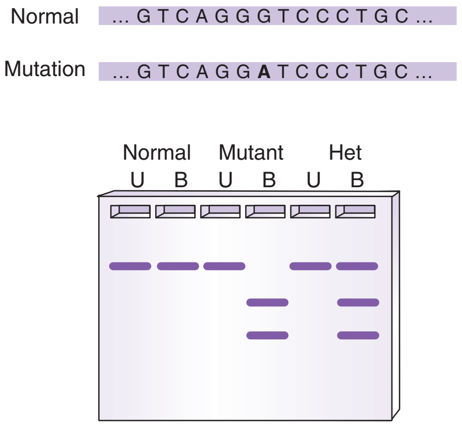

Normal ... G T C A G G G T C C C T G C ...
Mutation ... G T C A G G A T C C C T G C ...
Normal
Mutant
Het
U B U B U B

FIGURE 8.20 PCR-RFLP. The normal sequence (top line) is converted to a *Bam*H1 restriction site (GGATCC) by a G>A mutation. The presence of the mutation is detected by testing the PCR product with *Bam*H1. The bottom panel shows the predicted gel patterns for the homozygous normal, homozygous mutant, and heterozygous samples uncut (U) or cut with *Bam*H1 (B).

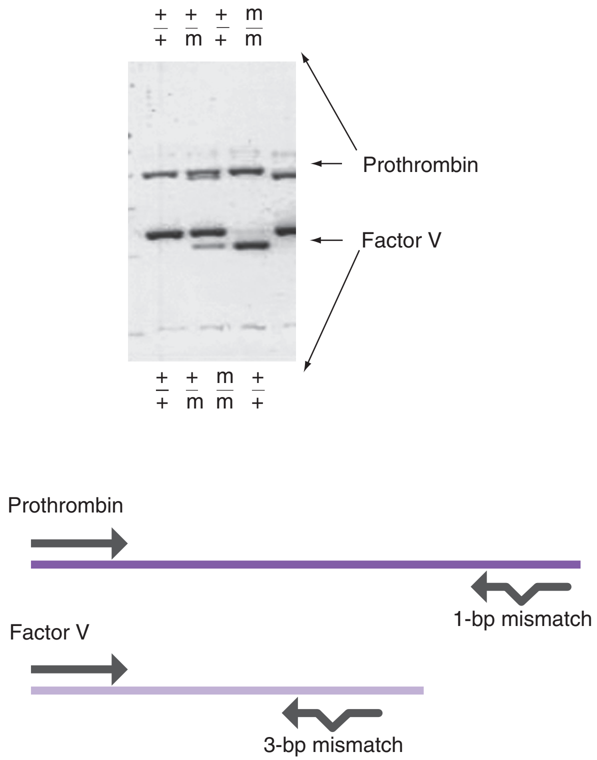


FIGURE 8.21 Multiplex PCR with mutagenic primers to detect mutations in factor V and prothrombin. The primer sequences are designed to generate a *Hind*III site in the PCR product if the mutations are present. The prothrombin and factor V PCR products are different sizes that can be resolved on the gel in a single lane.

are widely used for detection of commonly occurring mutations, such as *FLT3* kinase domain and HFE mutations.

PCR-RFLP can be multiplexed to detect more than one gene mutation simultaneously. This has been practical for the detection of separate gene mutations that affect the same phenotype, for example, factor V Leiden and prothrombin.[38] Alternatively, a combination of SSP-PCR and PCR-RFLP is also applied to simultaneous detection of mutations in more than one locus. An example is shown in Figure 8.21, in which a primer designed to produce a restriction site in the amplicon is used for each gene in a multiplex PCR. In the example, the primers are designed to generate a *Hind*III site in the amplicons. The PCR reaction and the *Hind*III digestion are performed in the same tube, and the products are separated on one lane of the gel.[39] This procedure is used in the clinical analysis of factor V Leiden and prothrombin mutations.

Nonisotopic RNase Cleavage Assay

Nonisotopic RNase cleavage assay (NIRCA) is a heteroduplex analysis using duplex RNA.[40,41] The sequences to be scanned are amplified using primers tailed with promoter sequences of 20 to 25 bp. T7 or SP6 phage RNA polymerase promoters are most often used for this purpose. Following amplification, the PCR products with the promoter sequences are used as templates for in vitro synthesis of RNA with the T7 or SP6 RNA polymerase enzymes. This reaction yields a large amount of double-stranded RNA (Fig. 8.22). The transcripts are denatured at 95°C and then renatured by cooling to room temperature. If a mutation is present, heteroduplexes form between normal and mutant transcripts. These mismatches in the RNA are targets for cleavage by RNase enzymes. A mixture of single-strand-specific *E. coli* RNase I and *Aspergillus* RNase T1 cleaves different types of mismatches. The remaining double-stranded RNA fragments can then be separated by agarose gel electrophoresis. As in DNA heteroduplex analysis, the size of the RNA fragments indicates the placement of the mutation. NIRCA has been applied to screening of several clinical targets, including factor IX,[7] *TP53*, *Jak2*, and *BRCA1*.

Cleavase Assay

The Cleavase assay is based on the characteristic enzymatic activity of a proprietary enzyme system (Cleavase).[42,43] Premixed reagents including Cleavase are added to a standard 96-well plate along with the test specimens (sample DNA or PCR products) and controls. Cleavase recognizes the structure formed by hybridization of the normal or mutant probes to the test sequences. During an isothermal incubation, if the probe and test sequence are complementary, two enzymatic cleavage reactions occur, ultimately resulting in a fluorescent signal (Fig. 8.23). The signal can be read by a standard fluorometer. The advantages of this method are the short hands-on time and optional PCR amplification. This method has been applied to several areas of clinical molecular diagnostics, including genetics, hemostasis, and infectious disease.

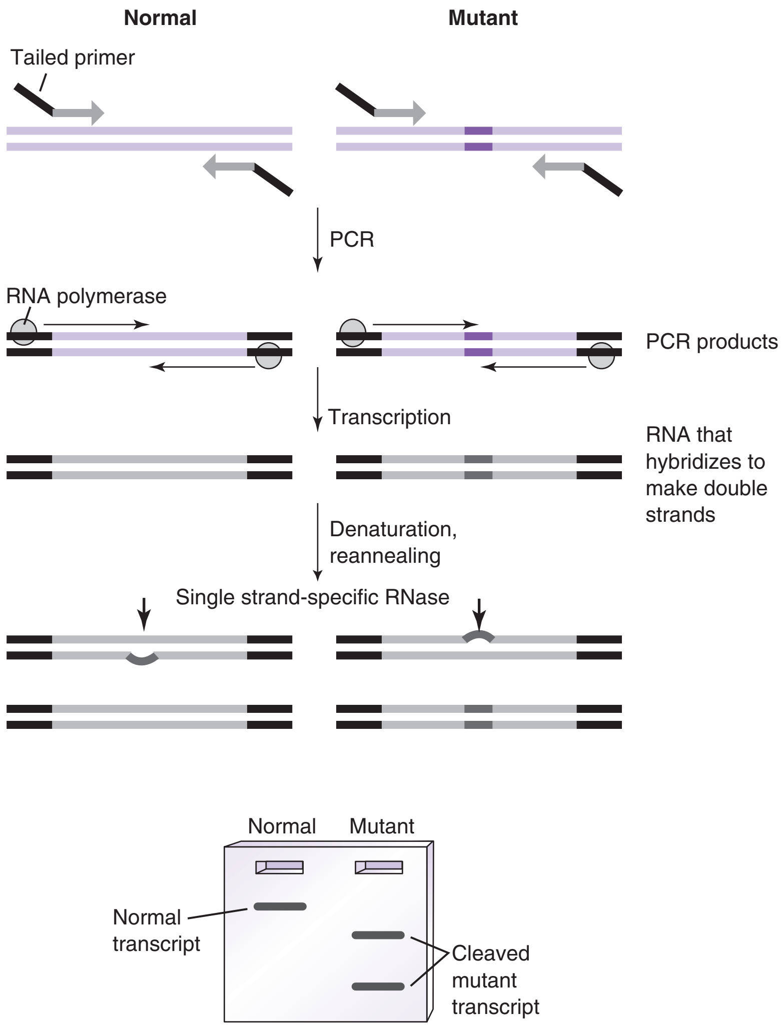

Normal
Mutant
Tailed primer
PCR
RNA polymerase
PCR products
Transcription
RNA that hybridizes to make double strands
Denaturation, reannealing
Single strand-specific RNase
Normal
Mutant
Normal transcript
Cleaved mutant transcript

FIGURE 8.22 NIRCA analysis. Normal *(left)* and mutant *(right)* transcription templates covering the area to be screened are produced by PCR with tailed primers carrying promoter sequences. RNA polymerase then transcribes the PCR products. The transcripts are denatured and reannealed, forming heteroduplexes between normal and mutant transcripts. RNase cleavage products can be resolved on native agarose gels.

Other Methods

The challenges of clinical laboratory requirements for robust, accurate, and sensitive assays have driven the discovery of new techniques and modification of existing techniques. As a consequence, many methods have been devised, especially for high-throughput screening. SSCP was a commonly used mutation screening method in clinical laboratories, but it became apparent from the use of SSCP that a single procedure may not be ideal for all genes—hence the development of diverse methods. Combinations of methods have also been proposed to increase sensitivity and detection, such as RFLP with modified primers. As the field of molecular diagnostics grows, the development of high-throughput methods has become a main focus. Array-based methods and massive parallel sequencing methods now provide the specific multiplex detection and sensitivity required for clinical applications. Instrument and reagent costs, which were relatively high for these technologies, are decreasing, especially relative to the information generated per test. Overall, the method selected will depend on the available instrumentation, the genetic target, and the nature of the mutation.

A summary of the methods described in this chapter is outlined in Table 8.2. The performance of each method

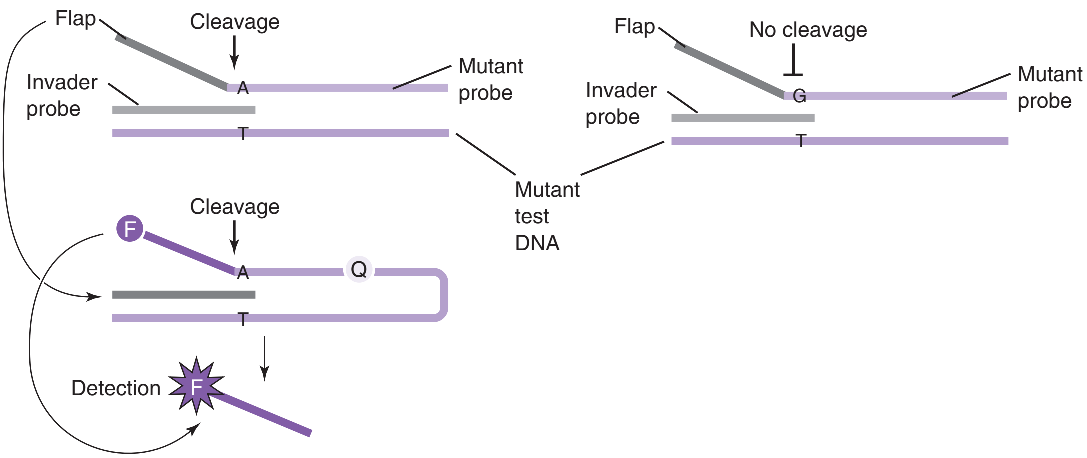


FIGURE 8.23 Cleavase single-color assay. Hybridization of supplied probe and anchor sequences to the input template *(upper left)* forms a structure that is the substrate for the cleavage enzyme. The enzyme removes the flap sequences, which form another hybridization structure with the labeled probe. The second cleavage reaction releases the fluorescent dye from the vicinity of the quencher on the probe, a fluorescent signal. If the template does not match the probe in the first hybridization *(upper right),* no cleavage occurs.

varies, depending on the specimen, template sequence, and type of mutation to be detected. For instance, methods that detect only mutations involving specific nucleotides can have 100% accuracy and specificity for these mutations but 0% for mutations affecting other nucleotides. Procedures that are developed by targeting specific mutations will perform for that target gene or region but may not work as well for other targets. For instance, hybridization methods may detect mutations in GC-rich sequence environments more accurately than in AT-rich sequences. Single-target methods are useful low-cost screening methods, but comprehensive and definitive methods such as direct sequencing are currently preferred, especially with complex diseases and new treatment strategies revealing growing numbers of clinically important variants.

GENE VARIANT NOMENCLATURE

Accurate testing and reporting of gene mutations require a descriptive and consistent system of expressing mutations and polymorphisms. Recommendations have been reported and generally accepted.[44] This section features general descriptive terms for basic alterations and structures.

For DNA and cDNA, the first nucleotide of the first amino acid in the sequence, usually A of ATG for methionine, is designated as position +1. The preceding nucleotide is position -1. There is no nucleotide position 0. Nucleotide changes are expressed as the position or nucleotide interval, the type of nucleotide change, the changed nucleotide, the symbol >, and finally the new nucleotide. For example, consider a nucleotide reference sequence: ATGCGTCACGAATTA. A substitution of a T for a C at position 7 in the DNA sequence (mutant sequence ATGCGTTACGAATTA) is expressed as c.7C>T; the "c." is for coding sequence. This format is intended to distinguish nucleotide changes from amino acid changes in proteins and is recommended for test reports. In large variant databases, such as those used for NGS analysis, however, the format c.C7T may be used.

A deletion of nucleotides 6 and 7, ATGCG__ACGAATTA, is expressed as c.6_7del or c.6_7delTC. An insertion of a TA between nucleotides 5 and 6, ATGCGTATCACGAATTA, is denoted c.5_6insTA. Duplications are a special type of insertion. For example, a duplication of nucleotides 4 and 5, ATGCGCGTCACGAATTA, is expressed as c.4_5dupCG. An insertion with a concomitant deletion, **indel,** has three alternate descriptive terms. For example, if TC at positions 6 and 7 of ATGCGTCACGAATTA is deleted from the reference sequence and GACA is inserted, the altered sequence, ATGCGGACAACGAATTA, is denoted as c.6_7delTCinsGACA, c.6_7delinsGACA or c.6_7>GACA. Inversion of nucleotides is designated by the nucleotides affected, "inv," and the number of nucleotides inverted. For example, inversion of GCGTCAC starting at position 3 to position 9 in the reference sequence (ATCACTGCGGAATTA) is c.3_9inv7.

Gene mutations in recessive diseases, where both alleles are affected, are indicated by the designation of each mutation separated by +. Thus, a 2357C>T mutation in one allele of a gene and a 2378delA mutation in the other allele on the homologous chromosome is written as c. [2357C>T]; [2378CdelA]. This is distinct from two mutations in the same allele, which is written as c.[2357C>T(;) 2378CdelA]. In cases of loss of heterozygosity, [0] indicates the absence of the entire reference coding sequence on the other chromosome. For example, c.[2357G>A];[0] denotes a G to A change at nucleotide 2357 on one chromosome and the loss of the gene on the homologous chromosome.

Mutations in introns of genomic DNA are indicated by the position of the mutation in the genomic sequence of the DNA or the position from the end of the coding sequence + the position in the intron. Thus, a G>T mutation 5 nucleotides into intron 2 that starts after the 91st nucleotide of exon 2 is designated as c.91+5G>T. If this same base change was at the 356th nucleotide in the genomic sequence, an alternative designation would be g.356G>C for genomic sequence.

At the protein level, numbering begins with the initial amino acid, methionine, in the protein sequence designated +1. The single-letter code has been used to convey protein sequence, but because of potential confusion with the single-letter designations, three-letter denotations are also acceptable. Stop codons are designated by X in either case. Amino acid changes are described by the amino acid changed, the position, and the new amino acid. Consider the protein sequence methionine–arginine–histidine–glutamic acid–leucine, or p.MRHEL. If the second amino acid, arginine (R), was substituted by serine (S), the mutation of the new amino acid sequence, p.MSHEL, would be p.R2S. A nonsense mutation in codon 4, mutant sequence MRHX, would be written as p.E4X, E4ter or E4*. Deletion of the arginine and histidine, M__EL, would be p.R2_H3del or p.R2_H3del2. Insertions are denoted by the amino acid interval, "ins," and the inserted amino acids. For instance, insertion of amino acids glycine (G), alanine (A), and threonine (T), making the altered amino acid sequence MRGATHEL, is indicated by p.R2_H3insGAT or, alternatively, p.R2_H3ins3. A short notation for frameshift mutations is the amino acid, its position, and "fs." A frameshift mutation affecting the histidine residue changing the amino acid sequence to MRCPLRGWX is simply p.H3fs, or p.H3fs* because frameshift mutations result in termination within a few amino acids. The length of the shifted open reading frame is indicated by adding X and the position of the termination codon. p.H3CfsX7 is a frameshift in codon 3 that changes a histidine to a cysteine and new reading frame ending in a stop at the seventh codon.

To distinguish between mutation nomenclature referring to genomic DNA, coding (complementary or copy) DNA, mitochondrial DNA, RNA, or protein sequences, the prefixes of "g.," "c.," "m.," "r.," and "p." are recommended, respectively. Furthermore, RNA sequences are written in lowercase letters. For example, c.89T>C in the coding DNA would be r.89u>c in RNA.

Complex changes and multiple concurrent mutations are reported as they occur. Some mutations, even with sequence information, cannot be positively determined and must be inferred, for example, additions or deletions of repeat units in repeated sequences. For these changes, it is assumed that the 3′-most repeat is the one affected, and the alteration is noted for that position. Updates and further clarifications of mutation nomenclature are continually being addressed. Current information and descriptors for more complex changes are available from the Human Genome Variation Society (HGVS) at http://www.HGVS.org/varnomen.

GENE NAMES

Gene nomenclature is different from gene names or sequence designation. The Human Genome Organization (HUGO) gene nomenclature committee has set rules for reporting or publishing gene names (see http://www.hugo-international.org/). Gene names should be capitalized and set in italics with no hyphens. Protein names are not italicized nor completely capitalized. For example, the *KRAS* gene codes for the K-Ras protein, and the *TP53* gene codes for the protein p53. Thus, a report will contain the official gene name and the appropriately named change in the DNA sequence, if present.

STUDY QUESTIONS

1. What characteristic of the genetic code facilitates identification of open reading frames in DNA sequences?

2. Compare and contrast EIA with western blots for the detection of protein targets.

3. On a size-exclusion column, large molecules will elute_______________ (before/after) small molecules.

4. MALDI methods separate ions by
   a. molecular volume.
   b. mass.
   c. charge.
   d. mass and charge.

5. What is a heteroduplex?

6. Exon 4 of the *HFE* gene from a patient suspected to have hereditary hemochromatosis was amplified by PCR. The G to A mutation, frequently found in hemochromatosis, creates an *Rsa*1 site in exon 4. When the PCR products are digested with *Rsa*1, what results (how many bands) would you expect to see if the patient has the mutation if no other *Rsa*1 sites are naturally present in the PCR product?

7. Which of the following methods identifies the presence of a mutation but not the mutant sequence?
   a. SSP-PCR
   b. SSCP
   c. PCR-RFLP
   d. NGS

8. What is the effect on the protein when a codon sequence is changed from TCT to TCC?

9. A reference sequence, ATGCCCTCTGGC, is mutated in malignant cells. The following mutations in this sequence have been described:

   ATGCGCTCTGGC
   ATGCCCTC - -GC
   ATAGCCTCTGGC
   ATGTCTCCCGGC

   Express these mutations using the accepted gene nomenclature (A = nucleotide position 1).

10. A reference peptide, MPSGCWR, is subject to inherited alterations. The following peptide sequences have been reported:

    MPSTGCWR
    MPSGX
    MPSGCWLVTGX
    MPSGR
    MPSGCWGCWR

    Express these mutations using the accepted nomenclature (M = amino acid position 1).

References

1. Mort M, Ivanov D, Cooper DN, Chuzhanova NA. A meta-analysis of nonsense mutations causing human genetic disease. *Human Mutation* 2008;29:1037–1047.
2. Zia A, Moses AM. Ranking insertion, deletion and nonsense mutations based on their effect on genetic information. *BMC Bioinformatics* 2011;12:1–14.
3. Yalow R, Berson SA. Immunoassay of endogenous plasma insulin in man. *Clinical Investigation* 1960;39:1157–1175.
4. Engvall E, Perlmann P. Enzyme-linked immunosorbent assay (ELISA). Quantitative assay of immunoglobulin G. *Immunohistochemistry* 1971;8:871–874.
5. Coons A. The beginnings of immunofluorescence. *Journal of Immunology* 1961;87:499–503.
6. Köhler G, Milstein C. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* 1975;256:495–497.
7. Schacht V, Kern JS. Basics of immunohistochemistry. *Journal of Investigative Dermatology* 2015;135:1–4.
8. Schertz F. The pure pigments, carotin and xanthophyll, and the Tswett adsorption method. *Plant Physiology* 1929;4:337–348.
9. Xiang Z, Tang AG, Ren YP, Zhou QX, Luo XB. Simultaneous determination of serum tryptophan metabolites in patients with systemic lupus erythematosus by high performance liquid chromatography with fluorescence detection. *Clinical Chemistry and Laboratory Medicine* 2010;48:513–517.

10. Wu N, Lippert JA, Lee ML. Practical aspects of ultrahigh pressure capillary liquid chromatography. *Journal of Chromatography A* 2001;911:1–12.

11. Kasumov T, Huang H, Chung YM, Zhang R, McCullough AJ, Kirwan JP. Quantification of ceramide species in biological samples by liquid chromatography-electrospray tandem mass spectrometry. *Analytical Biochemistry* 2010;401:154–161.

12. Skrzypczak M, Heinritz W, Schulz AM, Mierzejewski M, Froster UG. Evaluation of IP-RP-HPLC for length determination of the trinucleotide repeat fragments in Huntington's disease. *Journal of Chromatographic Science* 2010;48:55–58.

13. Zhu X, Wang K, Liu G, Wang Y, Xu J, Liu L, Li M, Shi J, Aa J, Yu L. Metabolic perturbation and potential markers in patients with esophageal cancer. *Gastroenterology Research* 2017;4:1–9. doi: 10.1155/2017/5469597

14. Tang H, Lu Y, Zhang L, Wu Z, Hou X, Xia H. Determination of volatile organic compounds exhaled by cell lines derived from hematological malignancies. *Bioscience Reports* 2017;37(3): BSR20170106.

15. Chace D. Mass spectrometry in newborn and metabolic screening: historical perspective and future directions. *Journal of Mass Spectrometry* 2009;44:163–170.

16. Orita M, Iwahana H, Kanazawa H, Hayashi K, Sekiya T. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proceedings of the National Academy of Sciences* 1989;86:2766–2770.

17. Schmalenberger A, Tebbe CC. Profiling the diversity of microbial communities with single-strand conformation polymorphism (SSCP). *Methods in Molecular Biology* 2014;1095:71–83.

18. Liu Y, Li Y, Zhao X. Imbalance of G and C contents influences the sensitivity of single-strand conformation polymorphism. *Molecular Biology Reports* 2010;37:1605–1609.

19. Soong R, Iacopetta BJ. A rapid nonisotopic method for the screening and sequencing of *p53* gene mutations in formalin-fixed, paraffin-embedded tumors. *Modern Pathology* 1997;10:252–258.

20. Sarkar G, Sommer SS. Screening for mutations by RNA single-strand conformation polymorphism (rSSCP): comparison with DNA-SSCP. *Nucleic Acids Research* 1992;20:871–878.

21. Bisceglia L, Grifa A, Zelante L, Gasparini P. Development of RNA-SSCP protocols for the identification and screening of CFTR mutations: identification of two new mutations. *Human Mutation* 1994;4:136–140.

22. Liu Q, Sommer SS. Restriction endonuclease fingerprinting (REF): a sensitive method for screening mutations in long contiguous segments of DNA. *BioTechniques* 1995;18:470–477.

23. Struewing J, Hartge P, Wacholder S, Baker SM, Berlin M, McAdams M, Timmerman MM, Brody LC, Tucker MA. The risk of cancer associated with the 185delAG and 5382insC mutations of BRCA1 and the 6174delT mutation of BRCA2 among Ashkenazi Jews. *New England Journal of Medicine* 1997;336:1401–1408.

24. Hussussian C, Struewing JP, Goldstein AM, Higgins PA, Ally DS, Sheahan MD, Clark WH, Tucker, MA, Dracopoli NC. Germline p16 mutations in familial melanoma. *Nature Genetics* 1994;8:15–21.

25. Dunn P. Novel approaches and technologies in molecular HLA typing. *Methods in Molecular Biology* 2015;1310:213–230.

26. Wittwer C, Gudrun HR, Gundry CN, Vandersteen JG, Pryor RJ. High-resolution genotyping by amplicon melting analysis using LC green. *Clinical Chemistry* 2003;49:853–860.

27. Reed G, Wittwer CT. Sensitivity and specificity of single-nucleotide polymorphism scanning by high-resolution melting analysis. *Clinical Chemistry* 2004;50:1748–1754.

28. Lyon E. Mutation detection using fluorescent hybridization probes and melting curve analysis. *Expert Reviews in Molecular Diagnostics* 2001;1:92–101.

29. Kutyavin I, Afonina IA, Mills A, Gorn VV, Lukhtanov EA, Belousov ES, Singer MJ, Walburger DK, Lokhov SG, Gall AA, Dempcy R, Reed MW, Meyer RB, Hedgpeth J. 3′-minor groove binder-DNA probes increase sequence specificity at PCR extension temperatures. *Nucleic Acids Research* 2000;28:655–661.

30. Moutereau S, Narwa R, Matheron C, Vongmany N, Simon E, Goossens M. An improved electronic microarray-based diagnostic assay for identification of MEFV mutations. *Human Mutation* 2004;23:621–628.

31. Lopez-Crapez E, Livache T, Marchand J, Grenier J. K-ras mutation detection by hybridization to a polypyrrole DNA chip. *Clinical Chemistry* 2001;47:186–194.

32. Fulton R, McDade RL, Smith PL, Kienker LJ, Kettman JR. Advanced multiplexed analysis with the FlowMetrix system. *Clinical Chemistry* 1997;43:1749–1756.

33. Ye S, Humphries S, Green F. Allele specific amplification by tetra-primer PCR. *Nucleic Acids Research* 1992;20:1152.

34. Oliphant A, Barker DL, Stuelpnagel JR, Chee MS. BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping. *BioTechniques* 2002;32:s56–s61.

35. Abramovitz M, Barwick BG, Willis S, Young B, Catzavelos C, Li Z, Kodani M, Tang W, Bouzyk M, Moreno CS, Leyland-Jones B. Molecular characterisation of formalin-fixed paraffin-embedded (FFPE) breast tumour specimens using a custom 512-gene breast cancer bead array-based platform. *British Journal of Cancer* 2011;105:1574–1581.

36. Louis M, Dekairelle AF, Gala JL. Rapid combined genotyping of factor V, prothrombin and methylenetetrahydrofolate reductase single nucleotide polymorphisms using minor groove binding DNA oligonucleotides (MGB probes) and real-time polymerase chain reaction. *Bioscience Reports* 2004;42:12.

37. Behrens M, Lange R. A highly reproducible and economically competitive SNP analysis of several well characterized human mutations. *Clinical Laboratory* 2004;50:305–316.

38. Baris I, Koksal V, Etlik O. Multiplex PCR-RFLP assay for detection of factor V Leiden and prothrombin G20210A. *Genetic Testing* 2004;8:381.

39. Huber S, McMaster KJ, Voelkerding KV. Analytical evaluation of primer engineered multiplex polymerase chain reaction–restriction fragment length polymorphism for detection of factor V Leiden and prothrombin G20210A. *Journal of Molecular Diagnostics* 2000;2:153–157.

40. Goldrick M, Kimball GR, Liu Q, Martin LA, Sommer SS, Tseng JY. NIRCA: a rapid robust method for screening for unknown point mutations. *BioTechniques* 1996;21:106–112.

41. Kambas K, Mitroulis I, Kourtzelis I, Chrysanthopoulou A, Speletas M, Ritis K. Fast and reliable mutation detection of the complete exon 11-15 JAK2 coding region using non-isotopic RNase cleavage assay (NIRCA). *European Journal of Haematology* 2009;83: 215–219.

42. Lyamichev V, Mast AL, Hall JG, Prudent JR, Kaiser MW, Takova T, Kwiatkowski RW, Sander TJ, de Arruda M, Arco DA, Neri BP, Brow MA. Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. *Nature Biotechnology* 1999;17:292–296.

43. Hall J, Eis PS, Law SM, Reynaldo LP, Prudent JR, Marshall DJ, Allawi HT, Mast AL, Dahlberg JE, Kwiatkowski RW, de Arruda M, Neri BP, Lyamichev VI. Sensitive detection of DNA polymorphisms by the serial invasive signal amplification reaction. *Proceedings of the National Academy of Sciences* 2000;97: 8272–8277.

44. den Dunnen J. Describing sequence variants using HGVS nomenclature. *Methods in Molecular Biology* 2017;1492:243–251.

Chapter 9

DNA Sequencing

Outline

- **DIRECT SEQUENCING**
    - *Manual Sequencing*
        - Chemical (Maxam–Gilbert) Sequencing
        - Dideoxy Chain Termination (Sanger) Sequencing
    - *Automated Fluorescent Sequencing*
        - Approaches to Automated Sanger Sequencing
        - The Sequencing Ladder
        - Electrophoresis
        - Sequence Interpretation
- **PYROSEQUENCING**
- **BISULFITE DNA SEQUENCING**
- **RNA SEQUENCING**
- **NEXT-GENERATION SEQUENCING**
    - *Gene Panels*
    - *NGS Library Preparation*
    - *Targeted Libraries*
    - *Sequencing Platforms*
    - *Sequence Quality*
    - *Filtering and Annotation*
- **BIOINFORMATICS**
- **THE HUMAN GENOME PROJECT**
    - *Variant Associations With Phenotype*
        - The Human Haplotype Mapping Project
        - The 1000 Genomes Project

Objectives

9.1 List the components and the molecular reactions that occur in chain termination sequencing.
9.2 Discuss the advantages of dye primer and dye terminator sequencing.
9.3 Derive a text DNA sequence from raw sequencing data.
9.4 Describe examples of alternative sequencing methods, such as pyrosequencing and next-generation sequencing (NGS).
9.5 Show different technical approaches to NGS and the two approaches used most in clinical applications.
9.6 Describe how NGS sequencing libraries are made.
9.7 Distinguish primer and probe-based enrichment.
9.8 Define *bioinformatics,* and describe electronic systems for the communication and application of sequence information.
9.9 Recount the events of the Human Genome Project.
9.10 Explain how variant databases were developed following completion of the Human Genome Project.

DNA sequence information (the order of nucleotides in the DNA molecule) is used in the medical laboratory for a variety of purposes, including detecting mutations, typing microorganisms, identifying human haplotypes, and designating polymorphisms. Treatment strategies including targeted therapies are now selected based on the results of these techniques.[1]

DIRECT SEQUENCING

The importance of knowing the order, or sequence, of nucleotides on the DNA chain was appreciated in the earliest days of molecular analysis. Elegant genetic experiments with microorganisms indirectly detected molecular changes at the nucleotide level using phenotypic characteristics, such as nutrient requirements.

Indirect methods of investigating nucleotide sequence differences are still in use today. Without knowing the nucleotide sequence of the targeted areas, the results from many of these methods would be difficult to interpret; in fact, some methods would not be useful at all. Direct determination of the nucleotide sequence, or DNA sequencing, is the most definitive molecular method to identify genetic lesions.

Manual Sequencing

Direct determination of the order, or sequence, of nucleotides in a DNA polymer is the most specific and direct method for identifying genetic lesions (mutations) or polymorphisms, especially when looking for changes affecting only one or two nucleotides. Two types of sequencing methods were concurrently developed in the 1970s: **Maxam–Gilbert sequencing**[2] and **Sanger sequencing.**[3]

Chemical (Maxam–Gilbert) Sequencing

The Maxam–Gilbert chemical sequencing method was developed by Allan M. Maxam and Walter Gilbert. Maxam–Gilbert sequencing required a double- or single-stranded version of the DNA region to be sequenced, with one end radioactively labeled.

For sequencing, the labeled fragment, or **template,** was aliquoted into four tubes. Each aliquot was treated with a different chemical with or without high salt (Fig. 9.1). Upon addition of a strong reducing agent, such as 10% piperidine, the single-stranded DNA would break at specific nucleotides (Table 9.1).

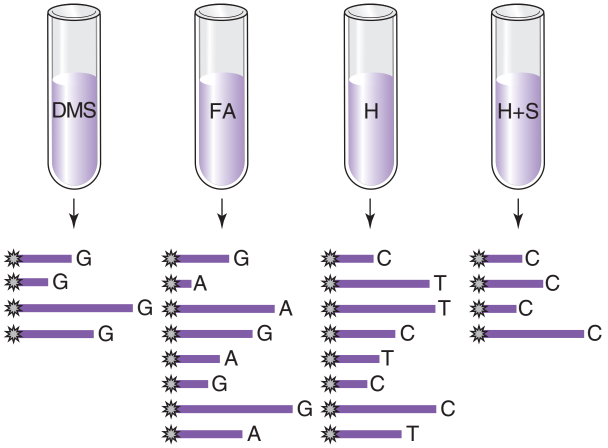


FIGURE 9.1 Chemical sequencing proceeds in four separate reactions in which the labeled DNA fragment is selectively broken at specific nucleotides. *DMS,* Dimethylsulphate; *FA,* formic acid; *H,* hydrazine; *H + S,* hydrazine + salt.

TABLE 9.1 Specific Base Reactions in Maxam–Gilbert Sequencing

Chain Breaks At:	Base Modifier	Reaction	Time (min at 25°C)
G	Dimethylsulphate	Methylates G	4
G + A	Formic acid	Protonates purines	5
T + C	Hydrazine	Splits pyrimidine rings	8
C	Hydrazine + salt	Splits only C rings	8

Advanced Concepts

To make a radioactive sequence template, (^{32}P)-ATP is added to the 5′ end of a DNA fragment, using polynucleotide kinase, or the 3′ end, using terminal transferase plus alkaline hydrolysis to remove excess adenylic acid residues. Double-stranded fragments labeled only at one end are also produced by using restriction enzymes to cleave a labeled fragment asymmetrically, and the cleaved products are isolated by gel electrophoresis. Alternatively, denatured single strands are labeled separately, or a “sticky” end of a restriction site is filled in, incorporating radioactive nucleotides with DNA polymerase.

After the reactions, the fragments were separated by size on a denaturing polyacrylamide gel. An example of Maxam–Gilbert sequencing results is shown in Figure 9.2. The sequence was inferred from the bands on the film. The lane in which that band appeared identified the nucleotide. Bands in the purine (G + A) or pyrimidine (C + T) lane were called based on whether they were also present in the G- or C-only lanes. In that way, the sequence was read from the bottom (5′ end of the DNA molecule) to the top (3′ end of the molecule) of the gel.

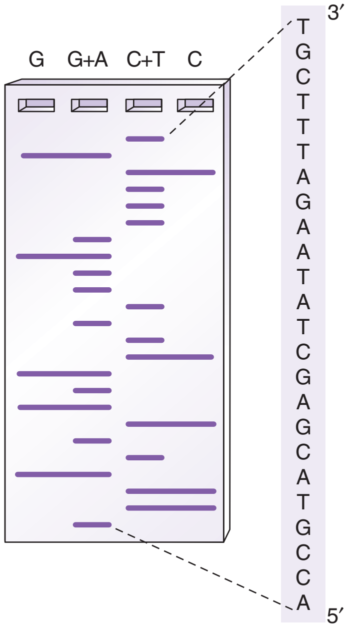


FIGURE 9.2 Products of a Maxam–Gilbert sequencing reaction. The gel is read from the bottom to the top. The size of the fragments gives the order of the nucleotides. The nucleotides are inferred from the lane in which each band appears. A or T is indicated by bands that appear in the G + A lane or C + T lane, respectively, but not in the G lane or the C lane. G is present in the G + A lane and the G lane. C is present in the C + T lane and the C lane.

Although Maxam–Gilbert sequencing was a relatively efficient way to determine short runs of sequence data, the method was not practical for high-throughput sequencing of long fragments. In addition, the hazardous chemicals hydrazine and piperidine required more elaborate precautions for use and storage. The method was therefore replaced by the **dideoxy chain termination** sequencing method for most sequencing applications.

Advanced Concepts

Polyacrylamide gels from 6% to 20% are used for sequencing. Bromophenol blue and xylene cyanol loading dyes are used to monitor the migration of the fragments. Run times range from 1 to 2 hours for short fragments (up to 50 base pairs [bp]) to 7 to 8 hours for longer fragments (more than 150 bp).

Dideoxy Chain Termination (Sanger) Sequencing

Dideoxy chain termination (Sanger) sequencing is a modification of the DNA replication process. A short, synthetic, single-stranded DNA fragment (primer) complementary to sequences just 5′ to the region of DNA to be sequenced is used for priming dideoxy sequencing reactions (Fig. 9.3). For detection of the products of the sequencing reaction, the primer is attached covalently at

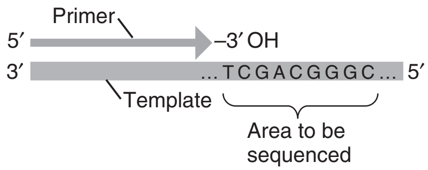


FIGURE 9.3 Manual dideoxy sequencing requires a single-stranded version of the fragment to be sequenced (template). Sequencing is primed with a short synthetic piece of DNA complementary to bases just before the region to be sequenced (primer). The sequence of the template will be determined by extension of the primer in the presence of dideoxynucleotides.

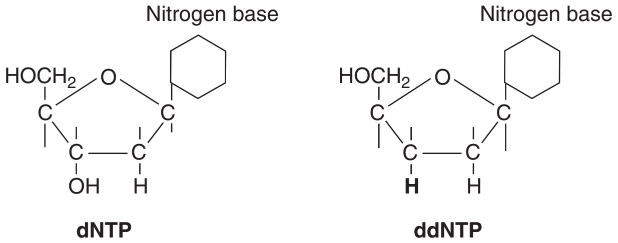


FIGURE 9.4 A dideoxynucleotide *(right)* lacks the hydroxyl group on the 3′ ribose carbon that is required for formation of a phosphodiester bond with the phosphate group of another nucleotide.

the 5′ end to a ^{32}P-labeled nucleotide or a fluorescent dye-labeled nucleotide. A previously used alternative detection strategy was to incorporate ^{32}P- or ^{35}S-labeled deoxynucleotides in the nucleotide sequencing reaction mix **(internal labeling).**

Just as in the in vivo DNA replication reaction, an in vitro DNA synthesis reaction would result in polymerization of deoxynucleotides to make full-length copies of the DNA template. For sequencing, modified **dideoxynucleotide (ddNTP)** derivatives are added to the reaction mixture. Dideoxynucleotides lack the hydroxyl group found on the 3′ ribose carbon of the deoxynucleotides (dNTPs; Fig. 9.4). DNA synthesis will stop upon incorporation of a ddNTP into the growing DNA chain (chain termination) because without the hydroxyl group at the 3′ sugar carbon, the 5′–3′ phosphodiester bond cannot be established to incorporate a subsequent nucleotide. The newly synthesized chain will terminate, therefore, with the ddNTP (Fig. 9.5).

Historical Highlights

The original dideoxy chain termination sequencing methods used in the late 1970s into the early 1980s required a single-stranded template. Templates up to a few thousand bases long were produced using M13 bacteriophage, a bacterial virus with a single-stranded DNA genome. This virus replicates by infecting *Escherichia coli,* in which the viral single-stranded circular genome is converted to a double-stranded plasmid, the **replication factor (RF).** The plasmid codes for viral gene products use the bacterial transcription and translation machinery to make new single-stranded genomes and viral proteins. To use M13 for template preparation, the RF was isolated from infected bacteria, cut with restriction enzymes, and the fragment to be sequenced was ligated into the RF. When the recombined RF was reintroduced into the host bacteria, M13 continued its life cycle producing new phages, some of which carried the inserted fragment. When the phages were spread on a lawn of host bacteria, plaques (clear spaces) of lysed bacteria formed by phage replication contained pure populations of recombinant phage. The single-stranded DNA was then isolated from the phage by picking plugs of agar from the plaques and isolating DNA from them.

Advanced Concepts

An advantage of the M13 template preparation method was that the primer that hybridizes to M13 sequences could be used to sequence any fragment ligated into the same site of the RF. Recombinant plasmids containing fragments to be sequenced include a short M13 region so that the **M13 universal primer** could still be used in some applications, even though the M13 method of template preparation is no longer practical.

For manual dideoxy sequencing, a 1:1 mixture of template and radioactively labeled primer is placed into four separate reaction tubes in sequencing buffer containing the sequencing enzyme and ingredients necessary for the polymerase activity (Fig. 9.6). Mixtures of all four dNTPs and one of the four ddNTPs are then added to each tube, with a different ddNTP in each of the four tubes.

Advanced Concepts

Polymerase chain reaction (PCR) products are currently used as sequencing templates. Residual components of the PCR reaction, especially

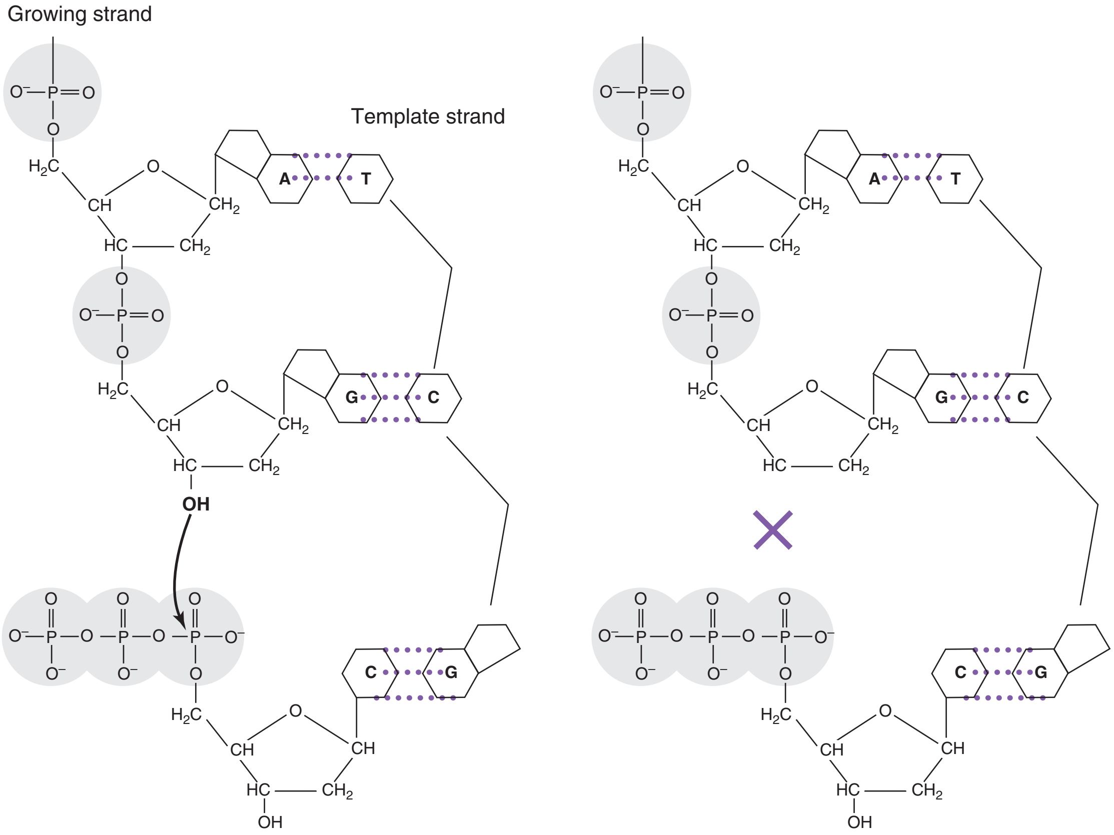


FIGURE 9.5 DNA replication *(left)* is terminated by the absence of the 3′ hydroxyl group on the dideoxyguanosine nucleotide (ddG, *right*). The resulting fragment ends in ddG.

primers and nucleotides, can interfere with the sequencing reaction and lower the quality of the sequencing ladder. PCR amplicons can be cleaned using solid-phase (column or bead) matrices, alcohol precipitation, or enzymatic digestion with alkaline phosphatase. Alternatively, amplicons can be run on an agarose gel and the bands eluted. The latter method provides not only a clean template but also confirmation of the product being sequenced. It is especially useful when the PCR reactions are not completely free of mis-primed bands or primer dimers.

The ratio of ddNTPs/dNTPs is critical for the generation of a readable sequence. If the concentration of ddNTPs is too high, polymerization will terminate too frequently early along the template. If the ddNTP concentration is too low, infrequent or no termination will occur. In the beginning days of sequencing, optimal ddNTP/dNTP ratios were determined empirically (by experimenting with different ratios). Sequencing reagent mixes have preoptimized nucleotide mixes.

With the addition of DNA polymerase enzyme to the four tubes, the reaction begins. After about 20 minutes, the reactions are terminated by addition of a stop buffer, which consists of 20 mM EDTA to chelate cations and stop enzyme activity, formamide to denature the products of the synthesis reaction, and gel loading dyes (bromophenol blue and/or xylene cyanol). All four reactions are carried out for equal times to provide consistent band intensities in all four lanes of the sequencing gel sequence.

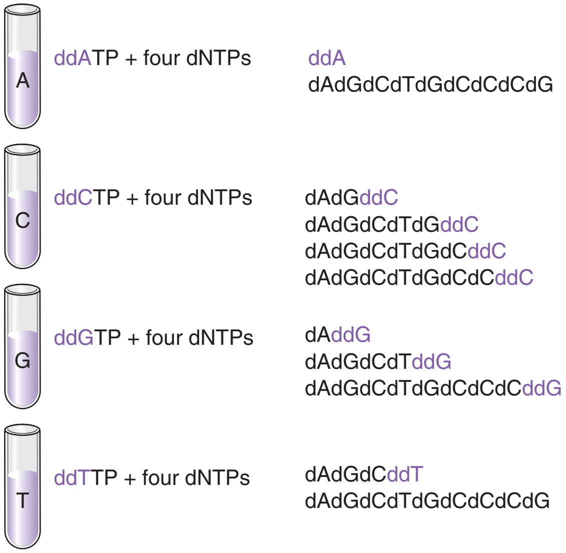


FIGURE 9.6 Components required for DNA synthesis (template, primer, enzyme, buffers, dNTPs) are mixed with a different ddNTP in each of four tubes (left). With the proper ratio of ddNTPs/dNTPs, the newly synthesized strands of DNA will terminate at each opportunity to incorporate a ddNTP. The resulting synthesis products are a series of fragments ending in either A (ddATP), C (ddCTP), G (ddGTP), or T (ddTTP). This collection of fragments is the sequencing ladder.

Advanced Concepts

Manganese (Mn^{++}) added to the sequencing reaction promotes equal incorporation of all dNTPs by the polymerase enzyme. Equal incorporation of the dNTPs makes for uniform band intensities on the sequencing gel, which eases interpretation of the sequence. Manganese increases the relative incorporation of ddNTPs as well, which will enhance the reading of the first part of the sequence by increasing intensity of the smaller bands on the gel. Modified nucleotides, deaza-dGTP and deoxyinosine triphosphate (dITP), are also added to sequencing reaction mixes to deter secondary structure in the synthesized fragments. Such additives as Mn^{++}, deaza-dGTP, and dITP are supplied in commercial sequencing buffers.

The sets of synthesized fragments are then loaded onto a denaturing polyacrylamide gel. The products of each of the four sequencing reactions are loaded into adjacent lanes, labeled A, C, G, or T, corresponding to the ddNTP in the four reaction tubes. Once the gel is dried and exposed to x-ray film, the fragment patterns are visualized by the signal on the ^{32}P-labeled primer (or incorporated deoxynucleotide). All fragments from a given tube will end in the same ddNTP; for example, all the fragments synthesized in the ddCTP tube end in C.

The four-lane gel electrophoresis pattern of the products of the four sequencing reactions is called a **sequencing ladder** (Fig. 9.7). The ladder is read to deduce the DNA sequence. From the bottom of the gel, the smallest (fastest-migrating) fragment is the one in which synthesis terminated closest to the primer. The identity of the ddNTP at a particular position is determined by the lane in which the band appears. If the smallest band is in the ddATP lane, then the first base is an A. The next larger fragment is the one that was terminated at the next position on the template. The lane that has the next larger band identifies the next nucleotide in the sequence. The sequence is thus read from the bottom (smallest, 5′-most) to the top (largest, 3′-most) fragments across or within lanes to determine the identity and order of nucleotides in the sequence.

Depending on the reagents and gel used, the number of bases per sequence read averages 300 to 400. Advances in enzyme and gel technology have increased this capability to over 500 bases per read. Sequencing reads are lengthened by loading the same ladders in intervals of 2 to 6 hours so that the larger bands are resolved with longer (e.g., 8-hour) migrations, whereas smaller bands will be resolved simultaneously in a 1- to 2-hour migration that was loaded 6 to 7 hours later.

As Sanger sequencing came into routine use, technology was improved significantly from these first manual sequencing procedures. Recombinant polymerase enzymes with in vitro removal of the exonuclease activity were faster and more processive (i.e., they stayed with the template longer, producing longer sequencing ladders). In addition, these engineered enzymes more efficiently incorporated ddNTPs and nucleotide analogs such as dITP or deaza-dGTP, which were used to deter secondary structure (internal folding and hybridization) in the template and sequencing products. Furthermore, sequencing was performed with double-stranded

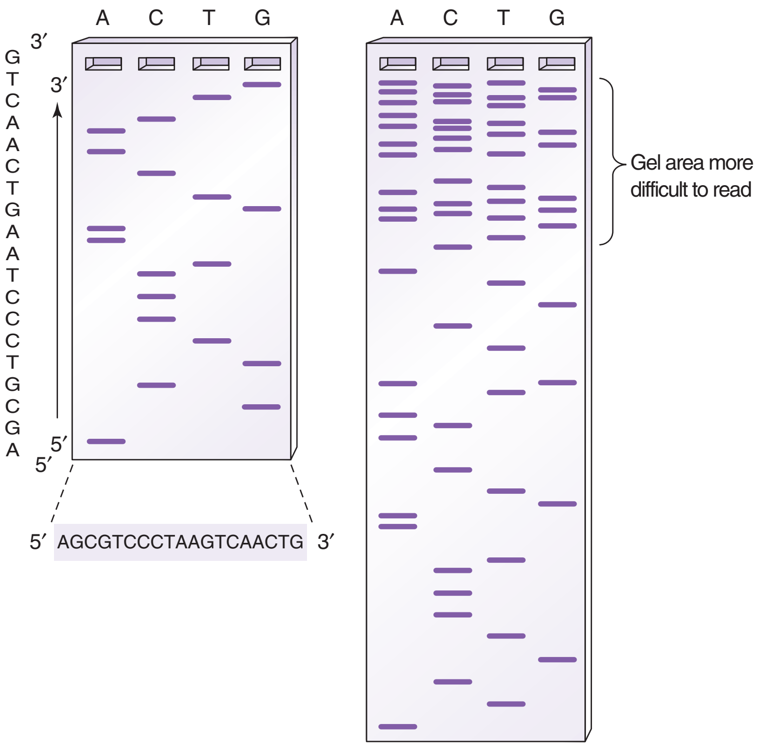


FIGURE 9.7 A sequencing ladder is read from the bottom of the gel to the top. The smallest (fastest-migrating) fragment represents the first nucleotide attached to the primer by the polymerase. Since that fragment is in lane A, from the reaction that contained ddATP *(left),* the sequence read begins with A. The next largest fragment is in lane T. The sequence, then, reads AT. The next largest fragment is in lane C, making the sequence ATC, and so forth up the gel. Larger bands on a sequencing gel can sometimes be compressed, limiting the length of sequence that can be read on a single gel run *(right).*

templates, eliminating the requirement for the preparation of single-stranded versions of the DNA to be sequenced.

Using heat-stable enzymes, the sequencing reaction took place in a thermal cycler **(cycle sequencing).** With cycle sequencing, timed manual starting and stopping of the sequencing reactions were not necessary. The labor savings in this regard increase the number of reactions that could be performed simultaneously; for example, a single operator could run 96 sequencing reactions (i.e., sequence 24 fragments) in a 96-well plate. Finally, improvements in fluorescent dye technology have led to the automation and throughput of the sequencing process and, more importantly, sequence determination.

Automated Fluorescent Sequencing

The chemistry for automated sequencing is the same as that described for manual sequencing, using double-stranded templates and cycle sequencing. Because cycle sequencing (unlike manual sequencing) does not require the sequential addition of reagents to start and stop the reaction, cycle sequencing was more easily adaptable to early high-throughput applications and automation. Universal systems combined automation of DNA isolation of the template and setup of the sequencing reactions.

Electrophoresis and reading of the sequencing ladder were also automated. A requirement for automated reading of the DNA sequence ladder is the use of fluorescent dyes instead of radioactive nucleotides to label the primers or sequencing fragments.

Advanced Concepts

Fluorescent dyes used for automated sequencing include fluorescein, rhodamine, and Bodipy (4,4-difluoro-4-bora-3a,4a-diaza-*s*-indacene) dye derivatives that are recognized by commercial detection systems.[4] Automated sequence readers excite the dyes with a laser and detect the emitted fluorescence at specific wavelengths. More advanced methods have been proposed to enhance the distinction between the dyes for more accurate determination of the sequence.[5]

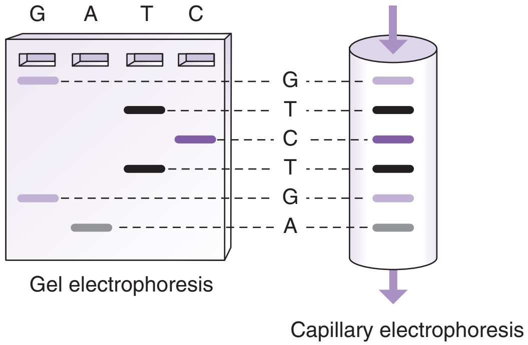


FIGURE 9.8 Instead of four gel lanes *(left)* fluorescent fragments can be run in a single gel lane or in a capillary *(right)*. Note that the sequence of nucleotides, AGTCTG, read by lane in the slab gel is read by color in the capillary.

Fluorescent dyes used for sequencing have distinct "colors," or peak wavelengths of fluorescence emission, that can be distinguished by automated sequencers. The advantage of having four distinct colors is that all four of the reaction mixes can be read in the same lane of a gel or on a capillary. Fluorescent dye color rather than lane placement will assign the fragments as ending in A, T, G, or C in the sequencing ladder (Fig. 9.8).

Approaches to Automated Sanger Sequencing

There are two approaches to automated fluorescent sequencing: **dye primer** and **dye terminator** sequencing (Fig. 9.9). The goal of both approaches is to label the fragments synthesized during the sequencing reaction according to their terminal ddNTP. Thus, fragments ending in ddATP, read as A in the sequence, will be labeled with a "green" dye; fragments ending in ddCTP, read as C in the sequence, will be labeled with a "blue" dye; fragments ending in ddGTP, read as G in the sequence, will be labeled with a "black" or "yellow" dye; and fragments ending in ddTTP, read as T in the sequence, will be labeled with a "red" dye. This facilitates reading of the sequence by the automated sequence.

In dye primer sequencing, the four different fluorescent dyes are attached to four separate aliquots of the primer. The dye molecules are attached covalently to the 5′ end of the primer during chemical synthesis, resulting in four versions of the same primer with different dye labels. The primer labeled with each "color" is added to four separate reaction tubes, one each with ddATP, ddCTP, ddGTP, or ddTTP, as shown in Figure 9.9. After addition of the remaining components of the sequencing reaction (see the previous section on manual sequencing) and of a heat-stable polymerase, the reaction is subjected to cycle sequencing in a thermal cycler. The products of the sequencing reaction are then labeled at the 5′ end, using the dye color associated with the ddNTP at the end of the fragment.

Dye terminator sequencing is performed with one of the four fluorescent dyes covalently attached to each of the ddNTPs instead of to the primer. The primer is unlabeled. A major advantage of this approach is that all four sequencing reactions are performed in the same tube (or well of a plate) instead of in four separate tubes. After addition of the rest of the reaction components and cycle sequencing, the product fragments are labeled at the 3′ end. As with dye primer sequencing, the "color" of the dye corresponds to the ddNTP that terminated the strand. Dye terminator sequencing has become the Sanger sequencing method of choice. The option of one reaction for all four nucleotides lowers the cost and labor of routine sequencing performed in many laboratories.

The Sequencing Ladder

After a sequencing reaction using fluorescent dye terminators, excess dye terminators are removed with columns or beads or by ethanol precipitation. Spin columns or bead systems bind the sequencing fragments to allow removal of residual sequencing components by rinsing with buffers. Alternately, the dye terminators are bound onto specially formulated magnetic beads, and the sequencing ladder is recovered from the supernatant as the beads are held by a magnet applied to the outside of the tube or plate.

The fragments of the sequencing ladder are completely denatured before running on a gel or capillary. Denaturing conditions (50°C to 60°C, formamide, urea denaturing gel) are maintained so that the fragments are resolved strictly according to size. Secondary structure affects migration speed and lowers the quality of the sequence. Before loading in a gel or capillary instrument, sequence ladders are cleaned, as described previously, to remove residual dye terminators, precipitated, and resuspended in formamide. The ladders are heated to 95°C to 98°C for 2 to 5 minutes and placed on ice just before loading.

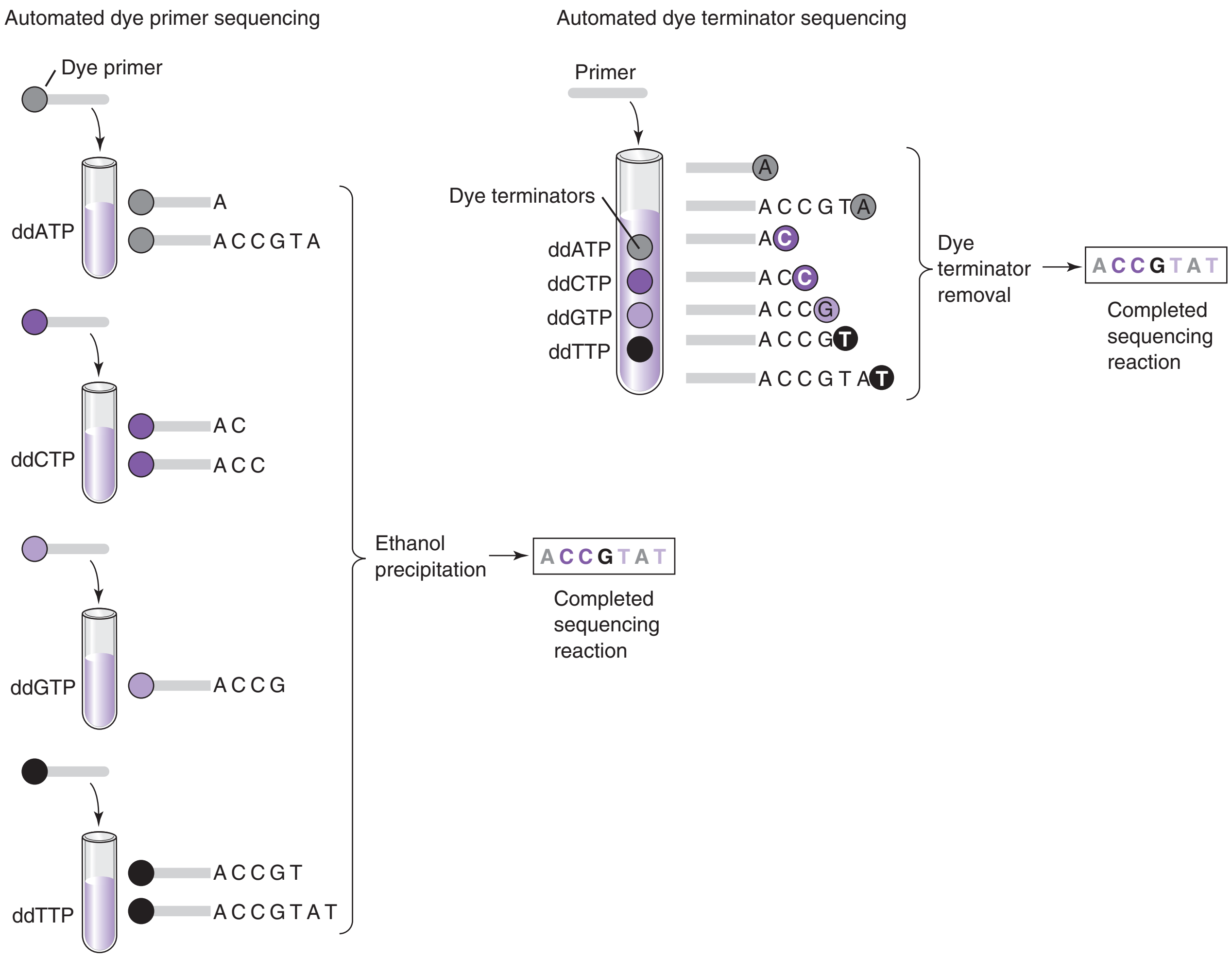

Automated dye primer sequencing
Dye primer
ddATP
A
A C C G T A
ddCTP
A C
A C C
ddGTP
A C C G
ddTTP
A C C G T
A C C G T A T
Ethanol precipitation
A C C G T A T
Completed sequencing reaction
Automated dye terminator sequencing
Primer
Dye terminators
ddATP
ddCTP
ddGTP
ddTTP
A
A C C G T A
A C
A C C
A C C G
A C C G T
A C C G T A T
Dye terminator removal
A C C G T A T
Completed sequencing reaction

FIGURE 9.9 Fluorescent sequencing chemistries. Dye primer sequencing uses labeled primers *(left)*. The reactions take place in separate tubes and the products of all four reactions are resolved together in one lane of a gel or in a capillary. Using dye terminators *(right)*, only one reaction tube is necessary because the fragments can be distinguished directly by the dideoxynucleotides on their 3′ ends.

Electrophoresis

The four sets of sequencing products in each reaction are loaded onto a single gel lane or capillary. The fluorescent dye colors, rather than lane assignment, distinguish which nucleotide is at the end of each fragment. Running all four reactions together not only increases throughput but also eliminates lane-to-lane migration variations that affect the accurate reading of the sequence. The migrating fragments pass a laser beam and a detector in the automated sequencer. The laser beam excites the dye attached to each fragment, causing the dye to emit fluorescence that is captured by the detector. The detector converts the fluorescence to an electrical signal that is imaged by computer software as a flash or peak of color.

Advanced Concepts

DNA sequences with high GC content are sometimes difficult to read due to intrastrand hybridization in the template DNA. Reagent preparations that include 7-deaza-dGTP (7-Deaza-2′-deoxyguanosine-5′-triphosphate) or dITP instead of standard dGTP improve the resolution of bands (peaks) in regions that exhibit GC band compressions, or bunching of peaks close together so that they are not resolved, followed by several peaks running farther apart.

Fluorescent detection equipment yields results as an **electropherogram,** rather than a gel pattern. Just as the gel sequence is read from the smallest (fastest-migrating) fragments to the largest, the sequencing software reads, or "calls," the bases from the smallest (fastest-migrating) fragments that first pass the detector to the largest based on the dye emission wavelength; that is, the software calls the base by the "color" of the fluorescence of the fragment as it passes the detector. The electropherogram is a series of peaks of the four fluorescent dyes as the bands of the sequencing ladder migrate by the detector. The software assigns one of four colors—red, black, blue, or green—associated with each of the fluorescent dyes and a text letter to the peaks for ease of interpretation.

As with manual sequencing, the ratio of ddNTPs/dNTPs is key to the length of the sequence read (how much of the template sequence can be determined). Too many ddNTPs will result in a short sequence read. Too low a concentration of ddNTPs will result in loss of sequence data close to the primer but give a longer read because the sequencing enzyme will polymerize further down the template before it incorporates a ddNTP into the growing chain. The quality of the sequence (height and separation of the peaks) improves away from the primer and begins to decline at the end. At least 400 to 500 bases can be easily read with most sequencing chemistries.

Sequence Interpretation

Base calling is the process of identification of bases in a sequence by sequencing software. It is analogous to the inspection of gel bands for quality, clarity, and separation. Interpretation of sequencing data from a dye terminator reaction depends on the quality of the electropherogram, which, in turn, depends on the quality of the template, the efficiency of the sequencing reaction, and the cleanliness of the sequencing ladder. Failure to clean the sequencing ladder properly results in bright flashes of fluorescence **(dye blobs)** that obliterate parts of the sequence read (Fig. 9.10). Poor starting material results in a poor-quality sequence that cannot be read accurately (Fig. 9.11). Clear, clean sequencing ladders are read accurately by the software, and a text sequence is generated. Sequencing software also shows the certainty of each base call in the sequence. When the base call is not clear, the letter N will replace A, C, T, or G. Less-than-optimal sequences are not accurately readable by software but may be readable by an experienced operator.

Software programs can compare two sequences or test sequences with reference sequences to identify mutations or polymorphisms. Regardless of whether a sequence variant (change from a reference sequence) is found, it is important to sequence both complementary strands of DNA to confirm sequence data. This is especially critical for confirmation of mutations or polymorphisms in a sequence (Fig. 9.12). Alterations affecting a single base pair may be subtle on an electropherogram, especially if the alteration is in the heterozygous form, or mixed with the normal reference sequence. Ideally, a genetically heterozygous mutation appears as two peaks of equal height but different colors directly on top of one another, that is, at the same position in the electropherogram. The overlapping peaks should be about half the height of single base peaks. Heterozygous deletions or insertions (e.g., the *BRCA* frameshift mutations) affect all positions of the sequence downstream of the mutation (Fig. 9.13). Somatic mutations in clinical specimens are sometimes difficult to detect because they may be diluted by normal sequences that mask the somatic change.

Several software programs have been written to interpret and apply sequence data from capillary electrophoresis. Software that collects the raw data from the instrument is supplied with the electrophoresis instruments. Software that interprets, compares, or otherwise manipulates sequence data is sometimes supplied with a purchased instrument or available online. A representative sample of these applications is shown in Table 9.2. Further sequence interpretation with regard to disease association and pathogenic significance requires the use of sequence databases and clinical trial information. This information is available from public websites and institutional "data commons" collections.

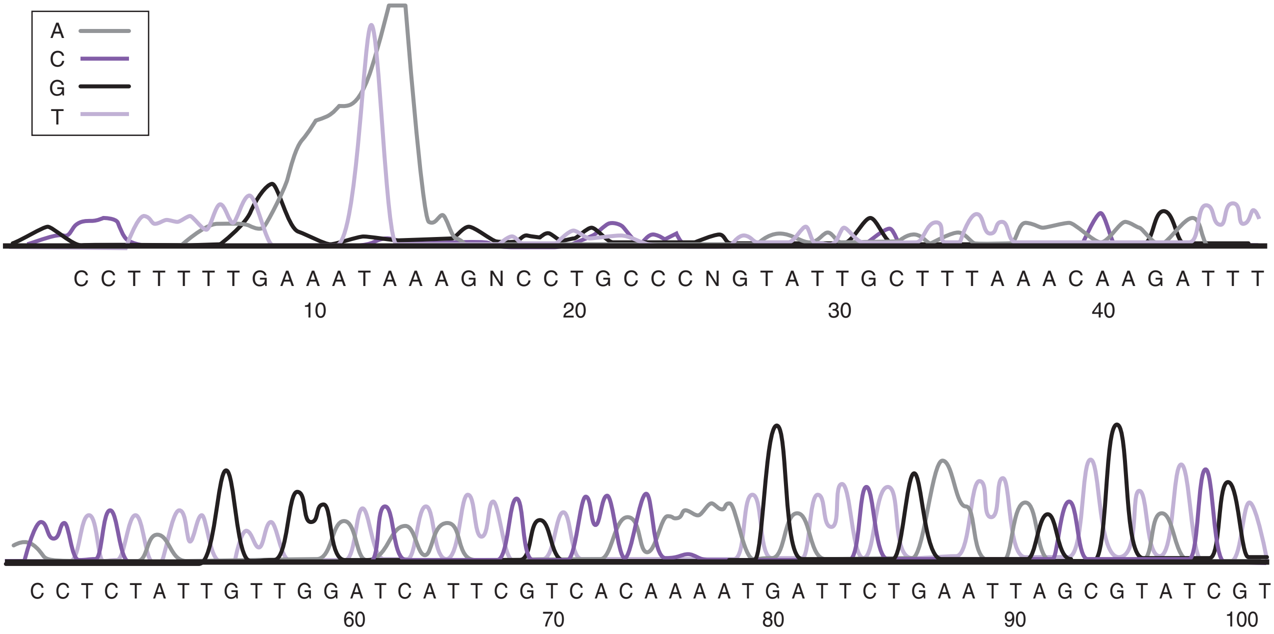


FIGURE 9.10 Electropherogram showing a dye blob at the beginning of a sequence (nucleotide positions 9 to 15). The sequence read around this area is not accurate. See Color Plate 5.

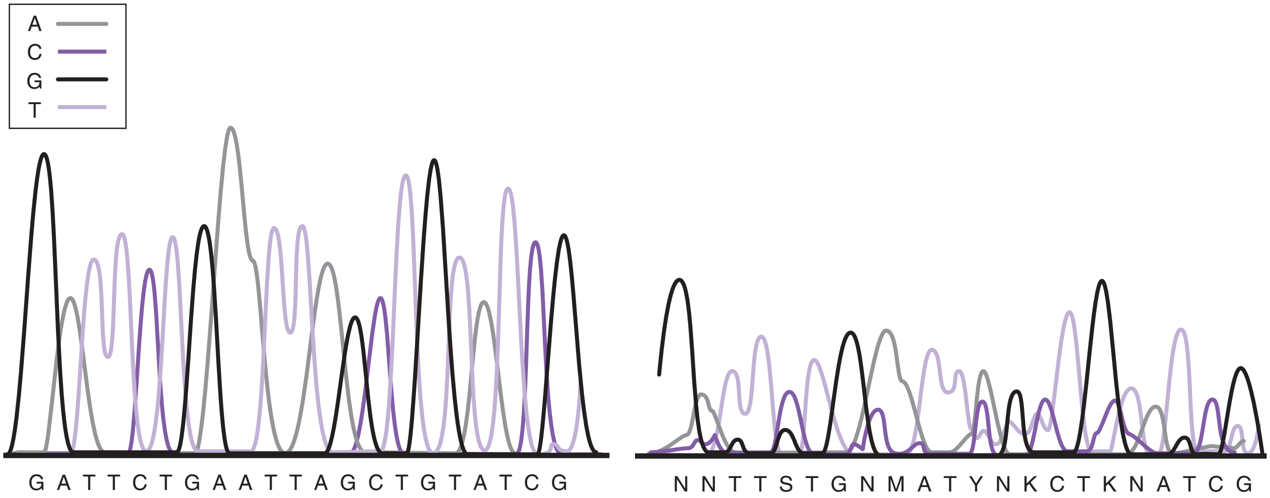


FIGURE 9.11 Examples of good sequence quality *(left)* and poor sequence quality *(right)*. Note the clean baseline on the good sequence; that is, only one color peak is present at each nucleotide position. Automatic sequence reading software will not accurately call a poor sequence. Compare the text sequences below the two scans. See Color Plate 6.

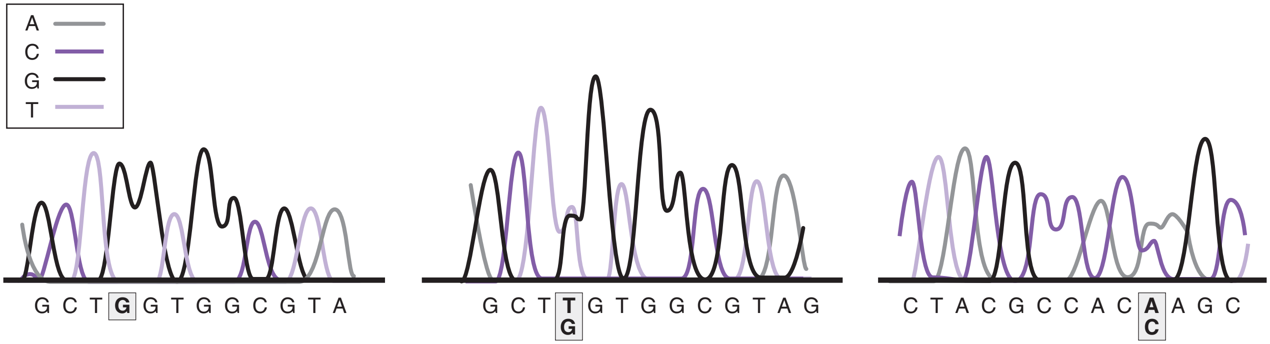


FIGURE 9.12 Sequencing of a heterozygous G to T mutation in exon 12 of the *KRAS* gene. The normal codon sequence is GGT *(left)*. The heterozygous mutation (GT, *center*) is confirmed in the reverse sequence (CA, *right*). See Color Plate 7.

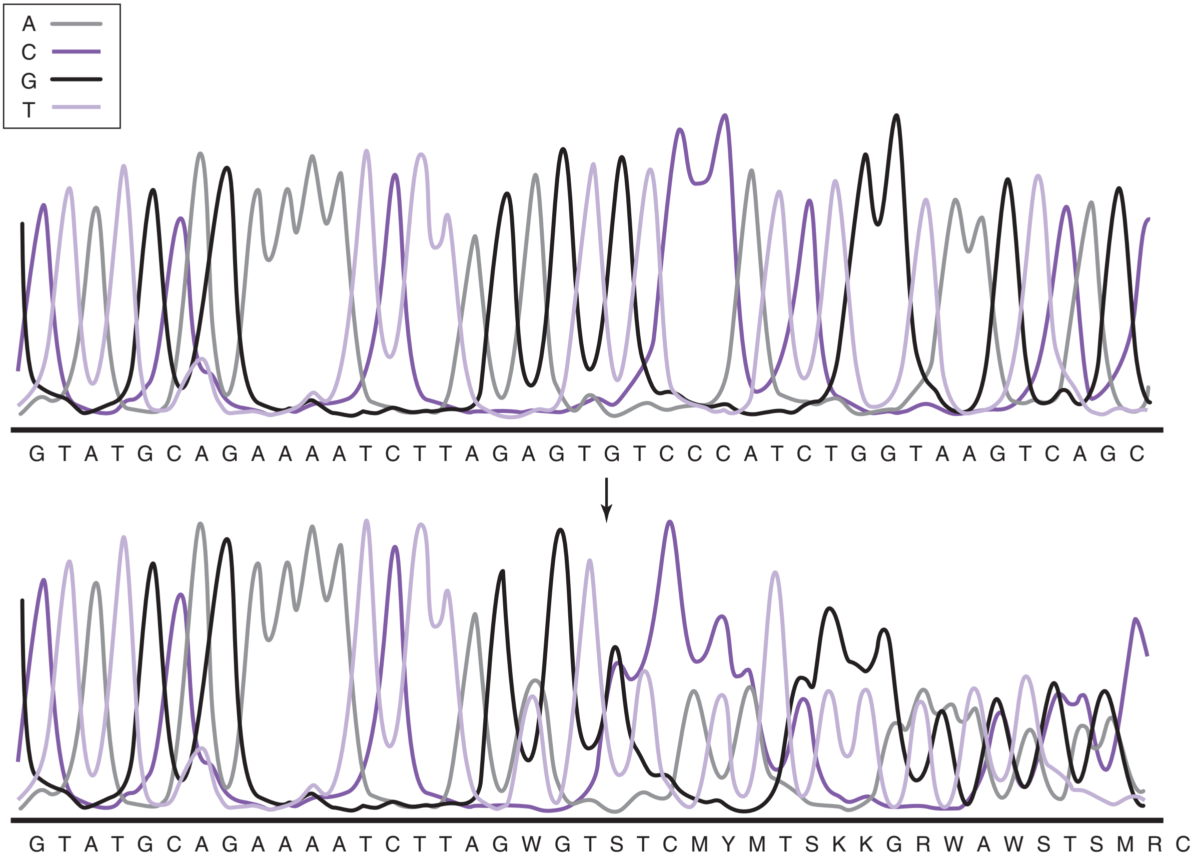


FIGURE 9.13 The 187 delAG mutation in the *BRCA*1 gene detected by Sanger sequencing. This heterozygous dinucleotide deletion is evident in the lower panel where, at the site of the mutation, two sequences are overlaid: the normal sequence and the normal sequence minus two bases. See Color Plate 8.

TABLE 9.2 Examples of Software Programs Used to Analyze and Apply Sequence Data

Software	Name	Application
BLAST	Basic Local Alignment Search Tool	Compares an input sequence with all sequences in a selected database
GRAIL	Gene Recognition and Assembly Internet Link	Finds gene-coding regions in DNA sequences
FASTA FASTQ	FAST-All derived from FAST-P (protein) and FAST-N (nucleotide) search algorithms Biological data with quality score	Rapidly aligns pairs of sequences by sequence patterns rather than individual nucleotides
Phred	Phred	Reads bases from original trace data and recalls the bases, assigning quality values to each base
Polyphred	Polyphred	Identifies single-nucleotide polymorphisms (SNPs) among the traces and assigns a rank indicating how well the trace at a site matches the expected pattern for an SNP
Phrap	Phragment Assembly Program	Uses user-supplied and internally computed data quality information to improve accuracy of assembly in the presence of repeats
TIGR Assembler	The Institute for Genomic Research	Developed by TIGR as an assembly tool to build a consensus sequence from smaller-sequence fragments
Factura	Factura	Identifies sequence features such as flanking vector sequences, restriction sites, and ambiguities
SeqScape	SeqScape	Provides mutation and SNP detection and analysis, pathogen subtyping, allele identification, and sequence confirmation
Assign	Assign	Identifies alleles for haplotyping
Matchmaker	Matchmaker	Identifies alleles for haplotyping

PYROSEQUENCING

Chain termination sequencing became the most widely used method to determine DNA sequence. Other methods were developed that yielded the same information but with less throughput capacity than the chain termination method. **Pyrosequencing** is an example of a method designed to determine a DNA sequence without having to make a sequencing ladder.[6,7] This procedure relies on the generation of light (luminescence) when nucleotides are added to a growing strand of DNA (Fig. 9.14). With this system, there are no gels, fluorescent dyes, or ddNTPs.

The pyrosequencing reaction mix consists of a single-stranded DNA template, sequencing primer, sulfurylase, and luciferase, plus the two substrates adenosine 5′ phosphosulfate (APS) and luciferin. One of the four dNTPs is added in a predetermined order to the reaction. If the nucleotide is complementary to the base in the template strand next to the 3′ end of the primer, DNA polymerase extends the primer. Pyrophosphate (PPi) is released with the formation of the phosphodiester bond between the dNTP and the primer. The PPi is converted to ATP by sulfurylase that is used to generate a luminescent signal by luciferase-catalyzed conversion of luciferin to oxyluciferin.

The process is repeated with each of the four nucleotides again added sequentially to the reaction. The generation of a signal indicates which nucleotide is the next correct base in the sequence. The results from a

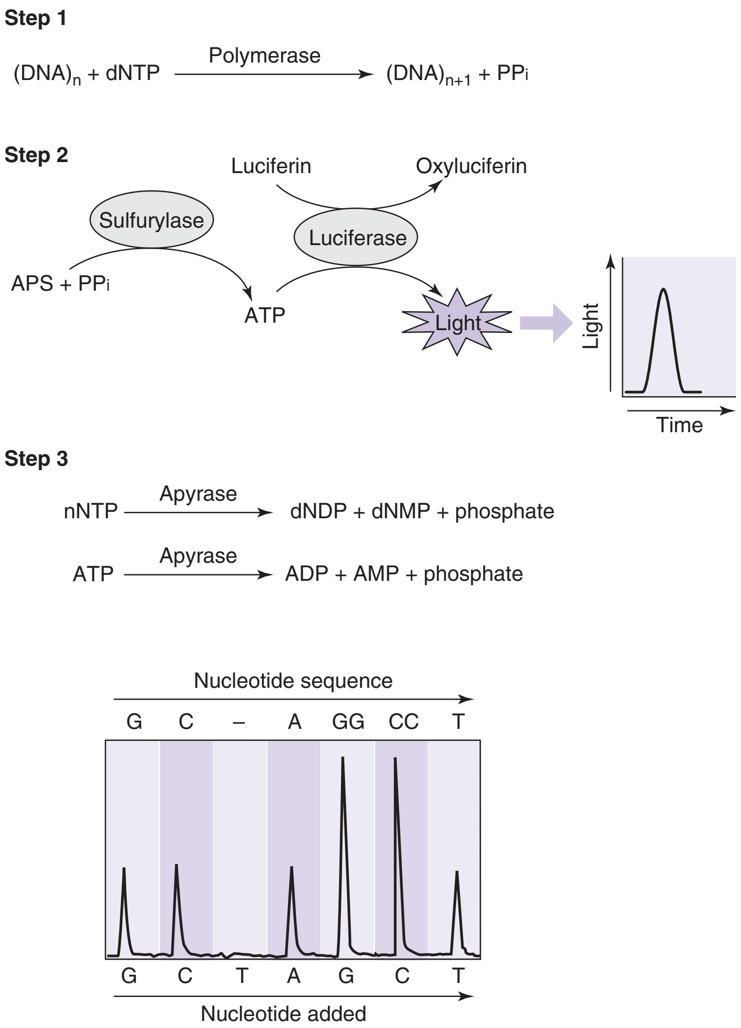

Step 1
(DNA)n + dNTP
Polymerase
(DNA)n+1 + PPi
Step 2
Sulfurylase
APS + PPi
ATP
Luciferin
Oxyluciferin
Luciferase
Light
Light
Time
Step 3
nNTP
Apyrase
dNDP + dNMP + phosphate
ATP
Apyrase
ADP + AMP + phosphate
Nucleotide sequence
G C – A GG CC T
G C T A G C T
Nucleotide added

FIGURE 9.14 Pyrosequencing is the analysis of pyrophosphate (PPi) released when a nucleotide base (dNTP) is incorporated into DNA *(top left)*. The released PPi is a cofactor for ATP generation from adenosine 5′ phosphosulfate (APS). Luciferase plus ATP converts luciferin to oxyluciferin with the production of light, which is detected by a luminometer. The system is regenerated with apyrase that degrades residual free dNTP and dATP (Step 3). As nucleotides are added to the system one at a time, the sequence is determined by which of the four nucleotides generates a light signal.

pyrosequencing reaction, a **pyrogram,** consist of peaks of luminescence associated with the addition of the complementary nucleotide (Fig. 9.14). If a sequence contains a repeated nucleotide, for instance, GCAGGCCT, the results would be dG peak, dC peak, dA peak, dG peak (double height), dC peak (double height), dT peak. The nucleotide sequence is called based on the order of nucleotide bases introduced to the sequencing reaction and the peak heights.

Pyrosequencing is most useful for short- to moderate-sequence analysis. It is therefore used mostly for detection of previously known mutation or single-nucleotide polymorphism (SNP) and typing (re-sequencing) rather than for generating new sequences. It has been used for applications in mutation detection,[8] infectious disease typing,[9,10] and DNA methylation analysis.[11]

BISULFITE DNA SEQUENCING

Bisulfite DNA sequencing, or methylation-specific sequencing, is chain termination sequencing designed to detect methylated cytosine nucleotides.[12] Methylation of cytosine residues to 5-methylcytosines in DNA is an important part of the regulation of gene expression and chromatin structure, affecting cell differentiation and diseases, including several types of cancer.

For bisulfite sequencing, 2 to 4 μg of genomic DNA is cut with restriction enzymes to facilitate denaturation. The enzymes should not cut within the region to be sequenced. The restriction digestion products are resolved on an agarose gel, and the fragments of the size of interest are purified from the gel. DNA from fixed tissue may be used directly without restriction digestion. The DNA is denatured with heat (97°C for 5 minutes) and exposed to bisulfite solution (sodium bisulfite, NaOH, and hydroquinone) for 16 to 20 hours. Buffer systems that protect DNA from bisulfite damage may be used to increase the yield of converted DNA. Overexposure to bisulfite can result in strand cleavage and loss of important regions of the DNA template. During the incubation with bisulfite, the cytosines in the reaction are deaminated, converting them to uracils, whereas the 5-methylcytosines are unchanged.

Advanced Concepts

Pyrosequencing requires a single-stranded sequencing template. Methods using streptavidin-conjugated beads have been devised to easily prepare the template. First the region of DNA to be sequenced is PCR-amplified with one of the PCR primers covalently attached to a biotin molecule. The double-stranded amplicons are then immobilized onto the beads and denatured with NaOH. After several washings to remove the non-biotinylated complementary strand (and all other reaction components), the sequencing primer is added and annealed to the pure single-stranded DNA template.

After the reaction, the treated DNA is cleaned, precipitated (or purified by adhering and washing on columns or beads), and resuspended for use as a template for PCR amplification. The primers used for amplification are altered to accommodate C to U changes in the primer-binding sites caused by the bisulfite treatment. For pyrosequencing, one primer is biotinylated for isolation of the single-stranded template.

The PCR amplicons are then sequenced by Sanger sequencing or pyrosequencing. Methylation is detected by comparing the treated sequence with the original sequence (before conversion) and noting where in the treated sequence cytosines are not changed to thymine (uracil); that is, the converted sequence will be altered relative to the reference sequence at the unmethylated C residues.

In Sanger sequencing, unmethylated cytosines will appear as red (thymine) instead of blue (cytosine) peaks on the electropherogram. In pyrosequencing, the relative light intensity of consecutive T and C additions to the reaction mix provide a quantitative degree of methylation. An example of pyrosequencing of bisulfite converted DNA is shown in Figure 9.15, where the color or height of the cytosine peaks relative to the thymine (uracil) peaks indicates the degree of methylation.

Detection methods other than sequencing have also been devised to detect DNA methylation, such as using methylation-sensitive restriction enzymes or enzymes with recognition sites generated or destroyed by the C to U changes. Other methods use PCR primers that will bind only to the converted or nonconverted sequences so that the presence or absence of PCR product indicates the methylation status. These methods, however, are not always applicable to the detection of methylation in unexplored sequences. As the role of methylation and epigenetics in human disease is increasingly recognized, bisulfite sequencing has become a popular method in the research laboratory. Clinical tests have been developed using this strategy as well.[13,14]

RNA SEQUENCING

The sequences of RNA transcripts are, for the most part, complementary to their DNA templates. Post-transcriptional processing of RNA, however, changes the RNA sequence relative to its encoding DNA. Alternative splicing and RNA editing may further modify the RNA sequence. Early methods to sequence RNA made use of ribonucleases to cut end-labeled RNA at specific nucleotides. Another approach was to infer mRNA sequence from amino acid sequence. The RNA transcript sequence can be determined from the sequencing of its complementary DNA; however, sequencing error may occur, mostly from the cDNA synthesis step.[15,16]

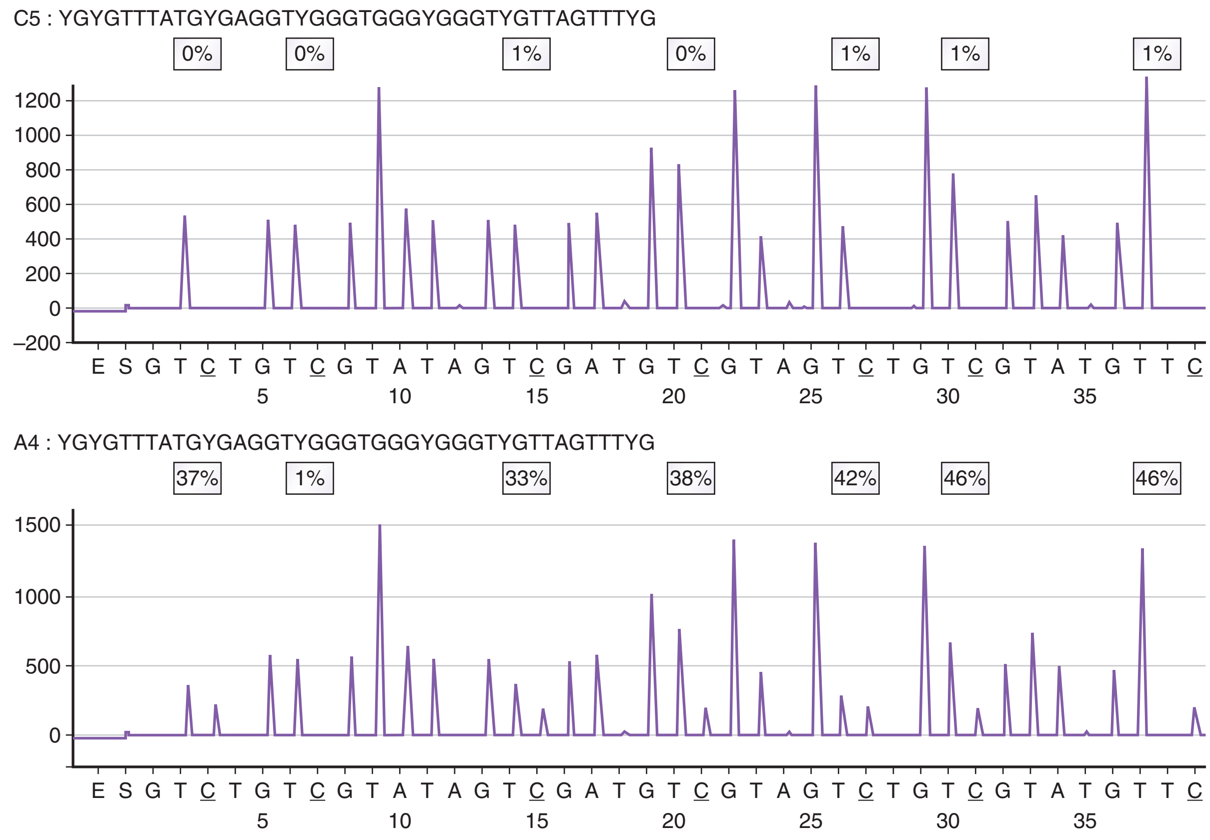


FIGURE 9.15 DNA methylation at cytosine residues detected by pyrosequencing of bisulfite-treated DNA. Exposure of a sequence to bisulfite will result in the conversion of unmethylated cytosines to uracils (T in the sequence). The pyrosequencing method will report the percent methylation that is the relative number of C to T nucleotides at each potentially methylated C position *(shaded)*. The C residues in the top panel are not methylated. All but one of the C residues in the bottom panel are methylated.

Direct sequencing of RNA has been proposed based on single-molecule sequencing technology and virtual terminator nucleotides.[17,18] In this method, mRNA is captured by immobilized polydT oligomers (Fig. 9.16). For those RNA species without polyA tails, an initial treatment with polyA polymerase is performed to add a 3′ A-tail. The 3′ ends of the captured RNA are chemically blocked to prevent extension in the sequencing step. Four reversibly dye-labeled nucleotides are then sequentially added. An image is taken, the extension inhibitors are cleaved, and alternating C, T, A, or G nucleotides are added, with imaging, cleavage, and rinsing between each nucleotide addition. After repeating this process many times (e.g., 120 cycles) the collected images are aligned and used to build the sequence from each poly(dT) anchor.

NEXT-GENERATION SEQUENCING

Data obtained from sequence analysis is best interpreted in context with population norms and variations; however, initially, few large sequence analyses were performed for multiple individuals. Furthermore, disease states involve a variety of sequence variants that can be important for diagnosis, prognosis, and treatment strategy. Although array studies were applied to this type of analysis, even the densest oligo array did not provide genomic-scale sequence data with single-base-pair resolution. Next-generation sequencing (NGS), also called massive parallel sequencing, was designed to sequence large numbers of templates carrying millions of bases simultaneously, in a run that takes a few hours. NGS

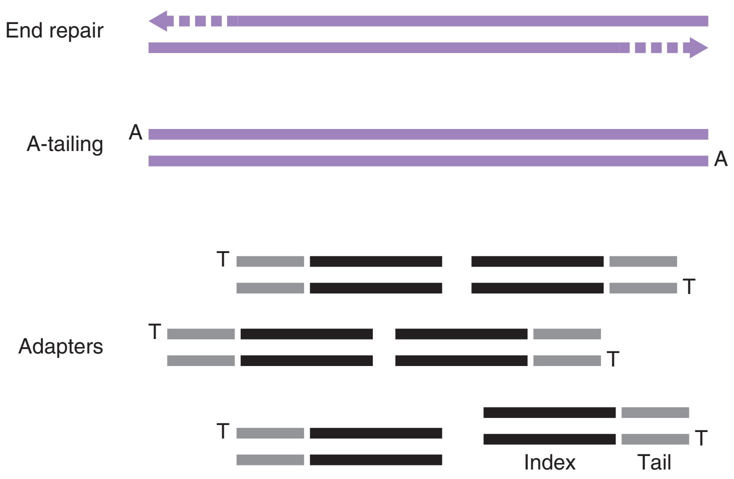


FIGURE 9.16 A next-generation sequencing library is created by the fragmentation of DNA. The fragmented DNA may have single-stranded ends that must be repaired back to double-stranded blunt ends by end repair. Addition of an A residue on the repaired blunt ends facilitates ligation of adapters carrying primer-binding sites for PCR amplification of the library.

technology has achieved gigabytes of sequencing data for a minimal cost, making genomic studies a routine component of both research and clinical analysis.

Historical Highlights

Early studies of DNA polymerase activity on an immobilized template led to the development of multiple template arrays that could be sequenced simultaneously. In 1997 high-throughput Selexa technologies were designed with capabilities of whole genome sequencing. High-throughput sequencing platforms developed in the mid-2000s resulted in a 50,000-fold drop in the cost of human genome sequencing from that of the Human Genome Project and led to the term *next-generation sequencing*. These technologies have increased in capacity and have been refined to address sequence complexity in genomes. The cost of sequencing a human genome has achieved the $1,000 cost point, which has expanded the use of sequencing analyses in the clinical laboratory.

NGS technologies include pyrosequencing, reversible dye terminator sequencing, ion-conductance sequencing, single-molecule sequencing, and sequencing by ligation. NGS requires novel methods of template preparation, such as emulsion PCR and bridge PCR, or single-molecule capabilities.[19] Powerful computer data assembly systems are required to organize the massive amounts of sequence information that are generated. These technologies can be used not only to sequence whole genomes but also to investigate populations of small genomes such as microbial diversity.[20]

Among the early challenges with massive sequencing was the integration of technologies without compromising accuracy or throughput.[21,22] These issues have been addressed with advances in bioinformatics and computer software. New challenges with system design, data accumulation and storage, clinical sensitivity, and data interpretation are being addressed, especially in dedicated sequencing facilities and commercial bioinformatics services.

NGS requires strong computer support as well as terabytes of storage space to accommodate large raw data sets. To prepare for NGS, clinical laboratories establish secure information channels and allocate space for preparation, loading, and operation of the sequencers. Interface with laboratory information systems and electronic medical records might also be arranged. Report templates are designed by the laboratory or commercial vendors and bioinformatics services.[23]

Two NGS technologies account for the majority of clinical sequencing applications: ion-conductance (pH)[24] and reversible dye terminator sequencing.[25] Both methods require the preparation of a sequencing library, sets of 100- to 500-bp-size fragments representing the regions to be sequenced. A library can represent a whole genome or a few specific gene regions where critical variants are likely to occur.

Gene Panels

The size and application of the sequencing library depend on the selection of genes to be sequenced or gene panels. Gene panels are probe or primer sets designed to amplify specific genes, regions, or entire exomes (all protein-coding sequences).[26] NGS might also be performed to compare sequences of many organisms (rRNA genes in microbial speciation) or to detect large numbers of possible base differences in a highly polymorphic gene such as *CFTR*. Gene panels have high technical sensitivity but require knowledge of the clinical diagnosis that would justify testing particular genes.

Gene panels have been designed for disease states, such as cardiomyopathies or muscular dystrophy or cancers. These panels range from a few (less than 20) target genes to more than a thousand target genes such as those used for solid organ cancers. "Hot-spot" panels target regions of specific genes known to affect treatment response, disease state, or clinical condition. Variants in these regions are referred to as "actionable" mutations; that is, a therapeutic or medical measure might be taken as a result of the presence of a variant. Targeted panels include critical genes in particular diseases such as hematological-cancer-specific panels for lymphoid or myeloid disorders or solid-tumor-specific panels for lung, colon, breast, or other cancers. Very large panels up to 3,000 genes or more provide a large amount of information for diagnostic, prognostic, and discovery purposes. These panels, however, may produce variants of unknown significance that must be assessed by pathologists and oncologists on a patient-specific basis. With the increase in novel treatment strategies, gene variants and combinations of gene variants previously not considered actionable can become so. Whole-exome sequencing is a method of gene discovery. This more challenging approach with regard to interpretation has proven beneficial in cases of suspected inherited gene variants.[27,28] Initially, beyond the scope of clinical analysis, whole-exome and even whole-genome sequencing have been increasingly incorporated in special cases. For routine clinical laboratory work, however, small- to medium-size 15- to 500-gene panels account for the majority of sequencing procedures.

NGS Library Preparation

A collection of DNA fragments to be sequenced is a sequencing **library.** Reversible dye terminator and ion-conductance sequencing are performed on DNA fragments less than 1,000 bp in length. Genomic DNA is fragmented by a number of methods, including shearing with high-frequency acoustic energy, sonication, nebulization (forcing DNA molecules through a small opening), or enzymatic treatments. Particular methods and how they are used (e.g., pressure levels used in nebulization) produce differently sized fragments (100 to 1,000 bp). The median fragment size can be checked by gel electrophoresis or microfluidics. Starting DNA concentrations and the DNA concentration of the library is best measured by fluorometry.

Advanced Concepts

Sequencing protocols and technologies differ with respect to the amount of required input genomic DNA. The lower limits range from 10 to 50 ng of DNA. For sequencing tumor DNA from fixed tissue, 140 mm^2 tissue with at least 30% tumor is recommended. Suboptimal amounts of starting DNA will compromise sequence quality and increase the risk of PCR artifacts. Fluorometric measurement of input (and library) concentrations is recommended over spectrometry to ensure the measurement of intact DNA.

Fragmented DNA produced by enzymatic or physical methods may be used directly for whole-exome or whole-genome sequencing. The fragments will have a mixture of 5′ and 3′ overhangs, some phosphorylated. To facilitate ligation to synthetic adaptors, single-stranded fragment ends are removed or filled in with nuclease or polymerase treatment. The 5′ ends are phosphorylated. The 3′ ends can be adenylated to further enable ligation to adapters with T overhangs (Fig. 9.16).

Adapters are synthetic short dsDNA pieces carrying sequences complementary to a single primer pair. The adaptors may also contain short sequences that will identify the sample (**indexing** or **bar coding**; Fig. 9.17). This allows analysis of multiple samples in the same reaction as the sequencing software will put together sequences

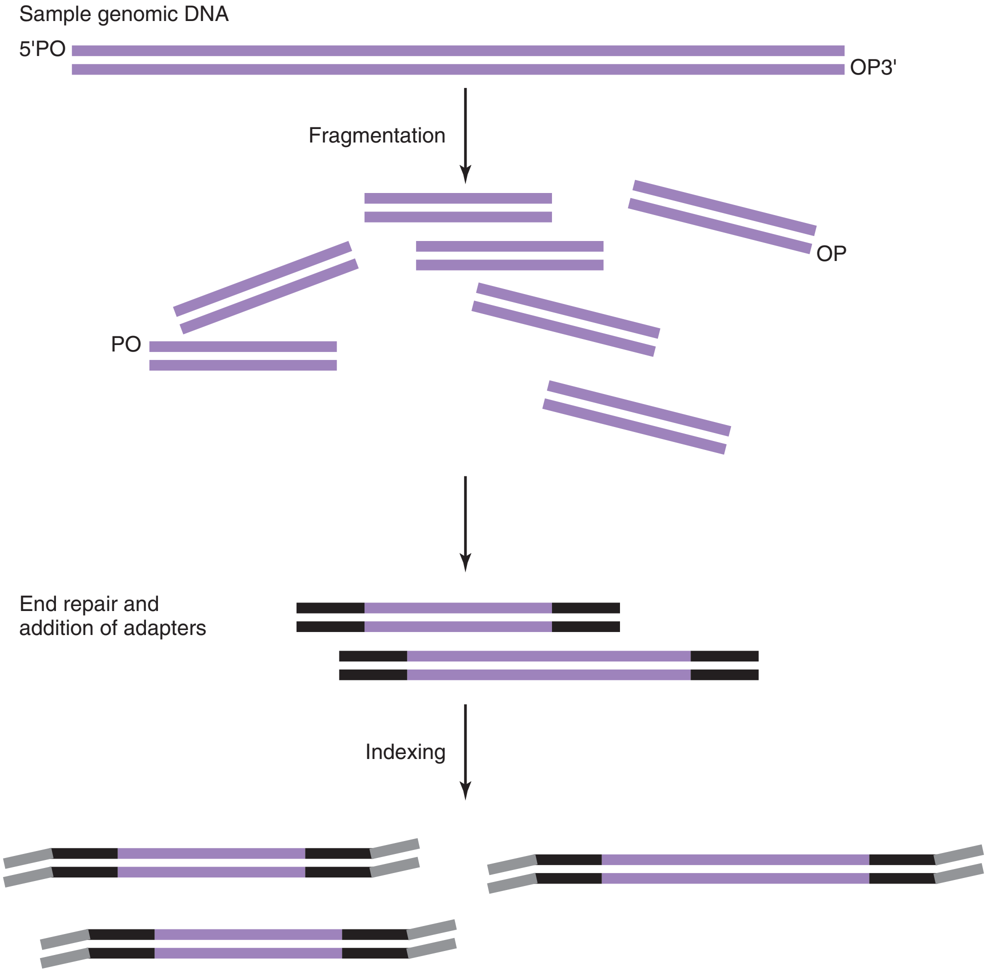


FIGURE 9.17 After fragmentation, end repair, and adapter ligation, bar codes or indexing may be performed by PCR amplification with tailed primers, or alternatively, the index sequences may be included in the adapters. Indexes are patient-specific so that multiple patient DNA can be sequenced in the same reaction and separated by their bar codes or indices after the sequencing is completed.

from fragments with the same bar code. Small genomes such as those of microbes or plasmids can be simultaneously fragmented and ligated to sequencing adapters in a single reaction tube. Reagent sets are commercially available for library preparation.

Targeted Libraries

Routine clinical sequencing of human DNA does not include the entire genome. Gene panels ranging from a few genes to whole exomes (all protein-coding regions) are employed, depending on the purpose for sequencing. The regions to be sequenced are enriched by probe hybridization or by amplification with region-specific primers.[29]

Probes are biotinylated oligonucleotides complementary to specific gene regions (Fig. 9.18). Targeted fragments to be sequenced are selected by hybridization with the biotinylated probe and captured with streptavidin-coated beads. The captured regions are ligated to adapters carrying primer-binding sites (or amplified with primer-binding sites included with short oligo probes) so that all reactions can proceed under the same amplification conditions in a single PCR reaction. Probe-based

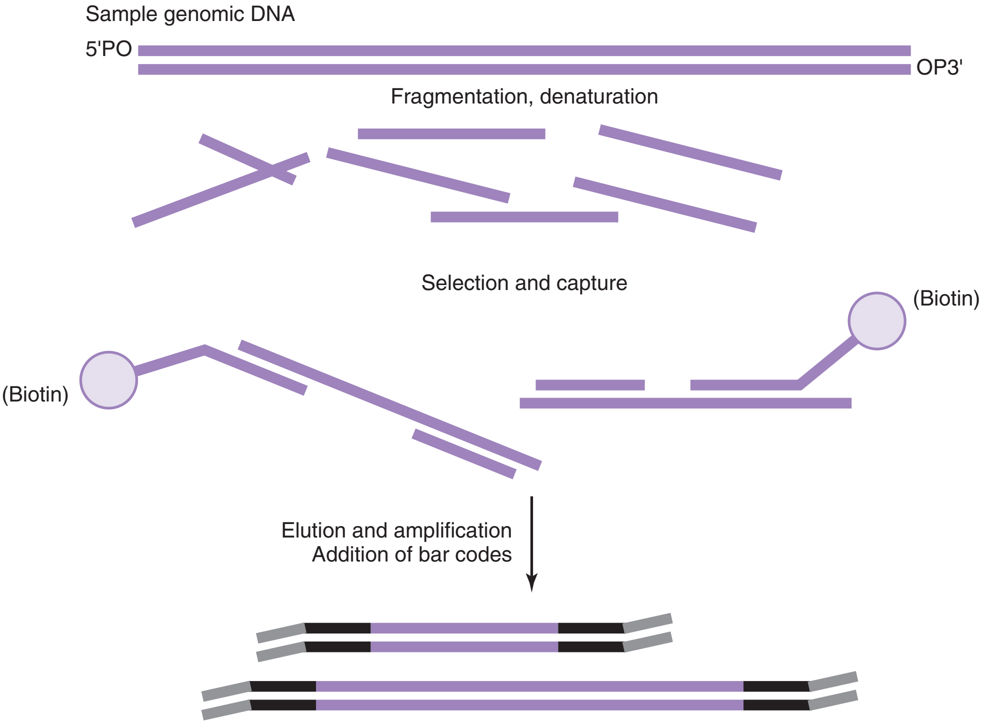


FIGURE 9.18 Targeted library preparation for NGS using probe enrichment. Fragmented DNA is denatured and hybridized to region-specific biotinylated probes. The probes are bead-captured, and the hybridized regions are amplified for sequencing. Probes may be short oligomers that can be extended across the region to be sequenced. The selected regions can then be amplified with tailed primers to add bar codes and sequencing primer-binding sites.

enrichment has the advantage of capturing sequences surrounding the region of interest and providing information from neighboring sequences. The presence of surrounding regions should be balanced because too much additional sequencing will affect the accurate sequencing of the targeted regions. The balance will depend on the average length of the DNA fragments.

Amplicon-based targeted libraries are selected by multiplex PCR with gene-specific primers tailed with binding sites for a secondary primer set (Fig. 9.19). After amplification, the secondary primers are tailed with index sequences that will identify (bar code or index) fragments from multiple samples in the same sequencing reaction and adapter sequences complementary to immobilized oligonucleotides anchored in the sequencing platform. These steps may be combined by tailing the initial multiplex PCR primers with the index and adapter sequences. Amplicon-based panel selection has the advantage of versatility and ease of use. Primer design is important, however, because sequence variations in the primer-binding sites may lower the efficiency of or even prevent amplification of particular fragments. Loss of library fragments from the sequenced regions, referred to as **allele dropout**, will cause inaccurate assessment of variant allele frequencies. Primers can be designed to produce overlapping sequences to cover less optimal regions. Paired-end or mate-pair primers produce coupled sequence fragments separated by 30 to 50 kb. By overlapping these reads, large variations not detectable in a few hundred base pairs such as translocations can be detected.

Both primer- and probe-based selections are affected by GC-rich sequencing targets. Secondary structure lowers the binding of primers and probes. GC-rich sequences also “clamp” primers in amplicon-based enrichment, lowering PCR efficiency. AT-rich regions may also be subject to poor hybridization, leading to loss of sequencing template fragments.

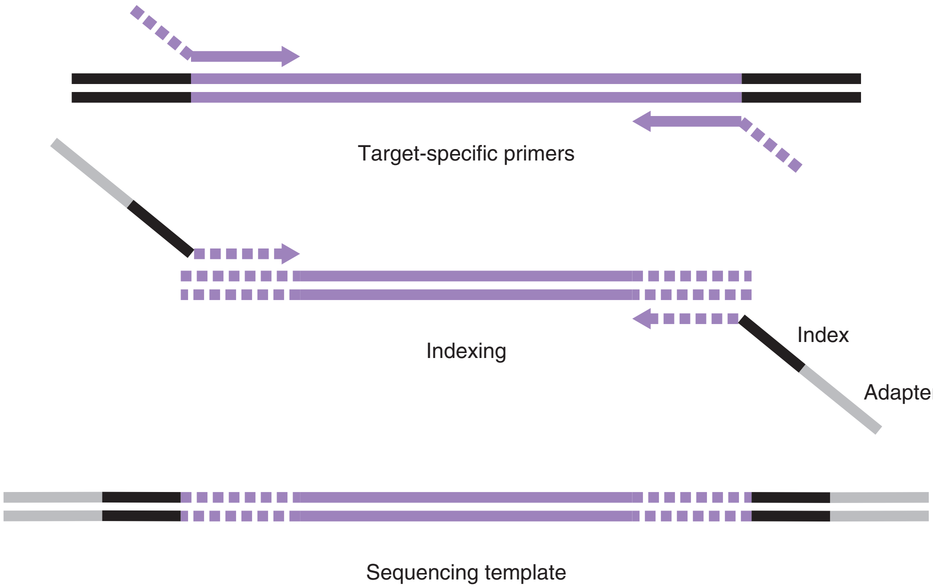


FIGURE 9.19 Targeted library preparation for NGS using amplification enrichment. Fragmented genomic DNA is end repaired and amplified with region-specific primers carrying binding sites for a single set of primers used in a second amplification. The second primer set has patient-specific index (bar-code) sequences.

Sequencing Platforms

After the introduction of NGS as a pyrosequencing technology, a variety of methods were developed for this purpose. The two most frequently used methods in clinical applications are ion-conductance and reversible dye terminator sequencing (Fig. 9.20). Both involve sequencing by synthesis and can be compared, chemically, to pyrosequencing and Sanger sequencing.

For ion-conductance sequencing, indexed libraries (gene panels) are amplified using primers immobilized on microparticles (beads) in an aqueous oil emulsion using adapters on the library fragments complementary to the immobilized primers (Fig. 9.20A). The beads carrying the amplicons (sequence templates) are placed on a solid surface (gene chip). The captured fragments are subjected to the addition of nucleotides in a predetermined order. If the nucleotide is complementary to the sequencing template, DNA polymerase will catalyze the formation of a phosphodiester bond. A hydrogen ion is released upon formation of the phosphodiester bond. The hydrogen ion will lower the pH of the reaction by a specific amount recorded by the sequencer (Fig. 9.21). This reaction occurs hundreds of thousands of times, producing sequence information from millions of sequencing panel library fragments.

In reversible dye terminator sequencing, captured or amplified fragments are hybridized to immobilized primers on a solid surface (flow cell). The fragments hybridize to the immobilized primers and are amplified by branch PCR into collections of products or **polonies** (Fig. 9.20B). Proper concentration (6 to 20 pMol) of the library DNA introduced to the flow cell will ensure that the polonies are evenly spaced on the flow cell. The polonies are sequenced in place by the sequential addition of fluorescently labeled nucleotides. If a nucleotide is complementary to the template next to the primer, DNA polymerase will extend the primer (form a phosphodiester bond). As in Sanger sequencing, each nucleotide is labeled with a specific color of fluor. An image is taken of the flow cell after each nucleotide addition (cycle), recording the presence of each added nucleotide color and location. After imaging, the fluorescent dyes are removed, and the next nucleotide is added (Fig. 9.22). Simultaneously, hundreds of thousands of polonies are sequenced in this way.

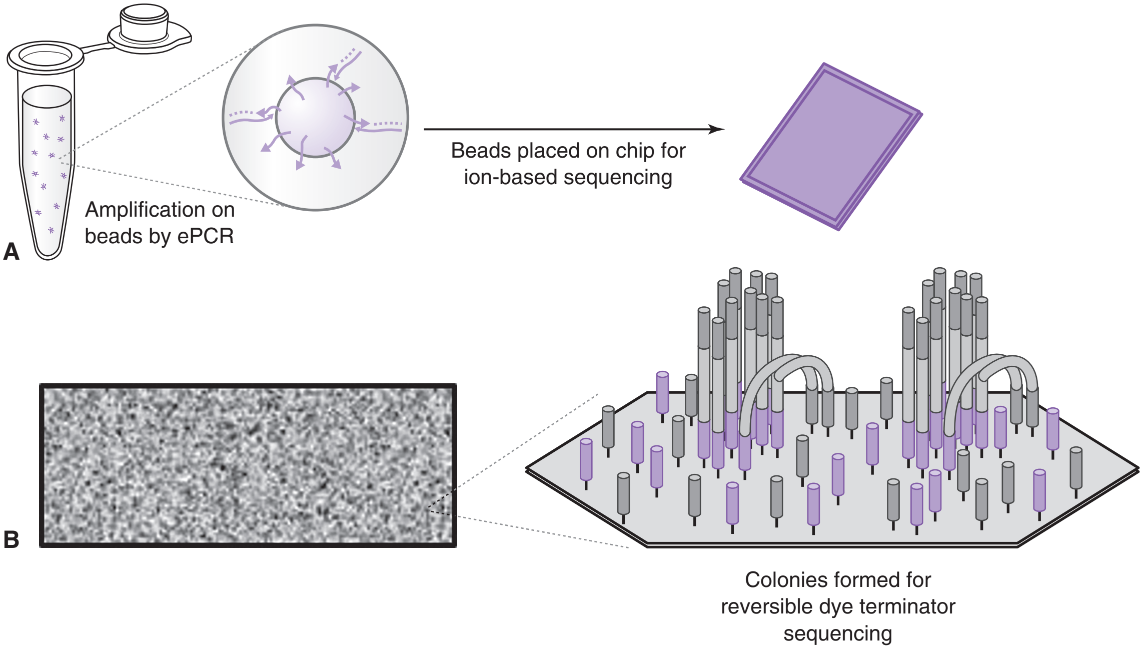

A
Amplification on beads by ePCR
Beads placed on chip for ion-based sequencing
B
Colonies formed for reversible dye terminator sequencing

FIGURE 9.20 (A) Library amplification for ion-conductance sequencing is performed in emulsion PCR. The bar-coded libraries prepared are amplified from primer-binding sites complementary to bead-immobilized primers. At the end of the ePCR reaction, the emulsion is broken and applied to a solid surface (chip) for sequencing. (B) For reversible dye terminator sequencing, the panel is amplified by bridge PCR through primer-binding sites complementary to primers immobilized on the flow cell. Amplification in place on the solid surface produces batches or polonies of sequencing templates distributed evenly across the flow cell.

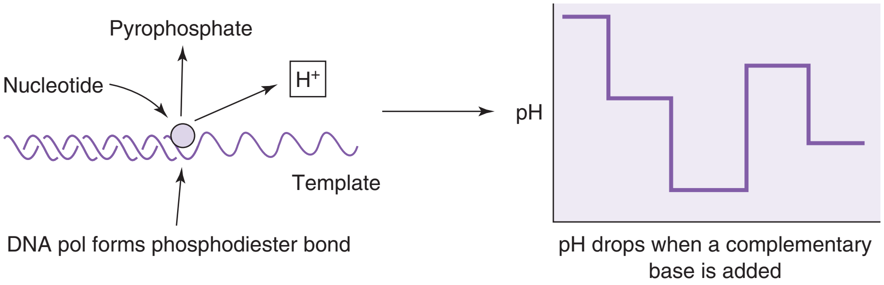

Pyrophosphate
Nucleotide
H^+
Template
DNA pol forms phosphodiester bond
pH
pH drops when a complementary base is added

FIGURE 9.21 In ion-conductance sequencing, when the nucleotide added to the reaction is complementary to the template, it is joined to the growing chain by DNA polymerase, releasing a hydrogen ion and drop in pH identifying that nucleotide. Sequencing software converts pH changes to the nucleotide sequence.

Both sequencing platforms are accurate and efficient, with comparable performance.[30] Proper controls include a no-template sequencing control and a reference sequence control. Sequence runs take from 2.5 hours to 2 days, depending on the platform and the size of the library being sequenced.

Other sequencing platforms such as sequencing by ligation[31] and nanopore sequencing[32] are used in research applications. Sequencing by ligation uses a pool of labeled oligonucleotide DNA ligase to identify the template sequence through the known probe sequences (Fig. 9.23). Nanopore sequencing has the advantage of

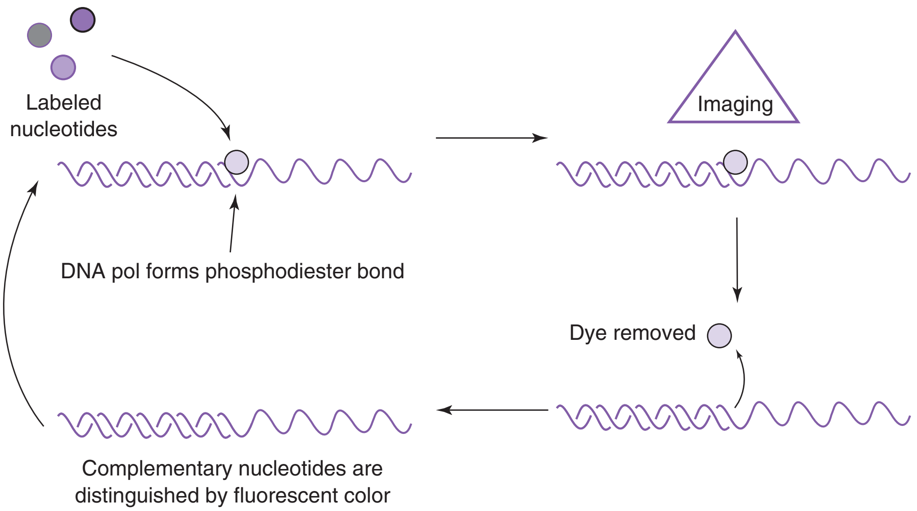

Labeled nucleotides
Imaging
DNA pol forms phosphodiester bond
Dye removed
Complementary nucleotides are distinguished by fluorescent color

FIGURE 9.22 In reversible dye terminator sequencing, labeled nucleotides are applied to the flow cell and incorporated into growing chains by DNA polymerase at each polony location. Images are taken after rounds of fluorescent nucleotide addition; the color at each polony location indicates the next nucleotide in that sequence. Once the image is taken, the fluorescent labels are removed. Following this, another round of nucleotides is introduced.

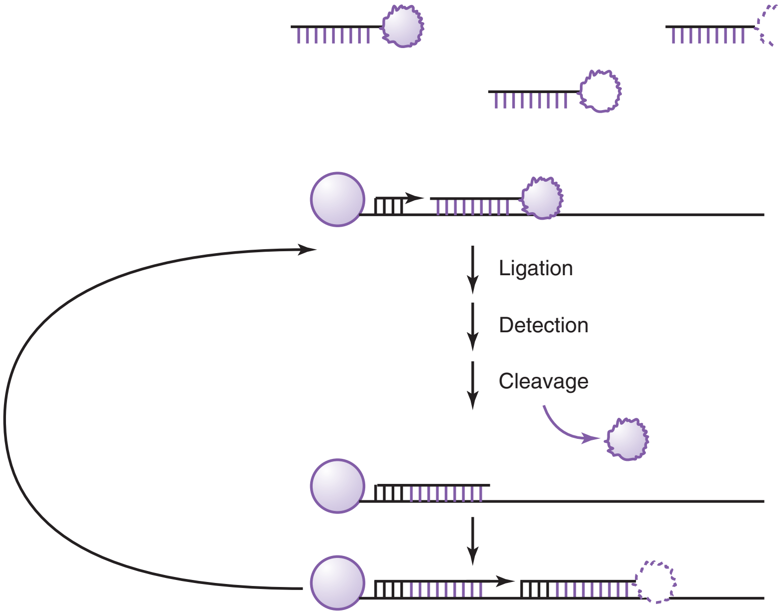

Ligation
Detection
Cleavage

FIGURE 9.23 Sequencing by ligation uses short fluorescently labeled oligomers that hybridize in short increments if they are complementary to the DNA template. The template DNA anchored to a glass slide is flooded with fluorescent-labeled oligonucleotides. If the oligo is complementary to the template, it is ligated, and then two bases are detected at a time. The oligonucleotide is cleaved, followed by the next round of ligation. Each time, two new nucleotides are detected.

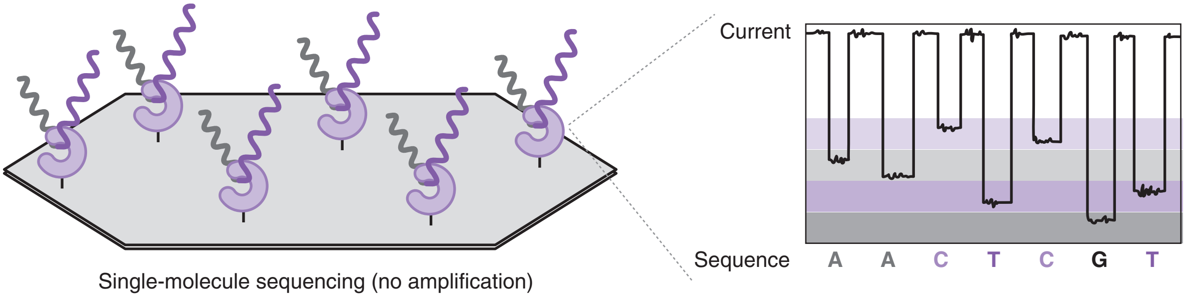


FIGURE 9.24 Long-read single-molecule sequencing uses protein ion channels through which one strand of each double-stranded DNA template is drawn. Each nucleotide passing through the pore changes the current in a characteristic way. This sequencing is rapid and does not require reassembly or short fragments for the final sequence.

not requiring fragmentation and amplification of the template DNA. One strand of long dsDNA molecules (up to 1 Mb) is drawn through protein pores. Each nucleotide is identified by a disruption in current as it passes through the pore (Fig. 9.24). This technology can also be used for direct RNA sequencing. Development of different technologies and improvement of existing technologies are actively occurring to further facilitate and widen the use of NGS.

Sequence Quality

Instrument collection and sequencing software will batch the sequences for each sample, based on the bar codes, and identify the nucleotide order in the process of base calling. Each base is assessed for quality of imaging (or conductance detection) and given a Phred score. Just as in Sanger sequencing, a Phred score of 2 to 3 (100- to 1,000-fold certainty of a correct call) is acceptable.

Each sequence is then compared to a reference sequence through read alignment. Reference sequences are considered "normal" in that there are no known significant variants; however, there is no real "normal" sequence, especially for human DNA. Variations from the reference may be the majority allele in the population, with the reference sequence carrying the minor allele. For human genome sequencing, reference genome hg19 was frequently used and reference genomes are updated periodically.[33] Reference sequences are free of known disease-related alleles, at least those found in the targeted panels.

The next step is variant identification based on comparison with the reference sequence. There are different types of variants, including single-nucleotide variants (SNVs), small insertion and/or deletion of nucleotides (indels), rearrangement of sequences (e.g., translocations), and copy-number variants (CNV; amplification or deletion of larger regions). Each of these types is handled differently by comparison software. Constitutional (genetically inherited) SNVs are identified in some programs based on a specific range of expected allele frequencies (variant allele/reference allele) for homozygosity or heterozygosity. Indels (up to 20 bp) can be identified by realignment, that is, multiple alignments (offset by one or more bases) that minimize base mismatches. Indels and even larger rearrangements can be detected by overlapping reads of paired end-primed sequences or by points of sequence diversions from 5′ and 3′ end reads (split-read analysis). Translocation breakpoints are often within introns or repetitive DNA sequences, or they contain overlaying sequence changes at the breakpoint, posing further challenges for variant identification. Optimal variant detection requires the use of the appropriate library primer design and software.

Once aligned, sequence variations from the reference (variants) are arranged in a variant call file (VCF). The VCF is a textual file that may be archived for further reference. Every variant is not of biological or clinical consequence. Some variants are synonymous or silent with regard to protein sequences. Others are common polymorphisms found in the population. Therefore, **annotation** are performed to identify critical variants.

TABLE 9.3 Annotation of Sequence Variants

Data	Description
Location of variant	Chromosome number, genomic coordinate (hg19 build)
Variant change	Reference allele, alternate allele detected in sample
Genetic state	Heterozygous, homozygous alternate
Quality of variant call	Quality/confidence score, sequencing depth at variant site (number of reads of variant and reference), probability of the reference or variant reads being balanced between + and − strands
Allele burden	The fraction of reads supporting the alternate allele (expect germline heterozygous alleles to be close to 0.5)
Variant type	The type of allele, either SNP, MNP, ins, del, or complex
Genomic position	Exon, intron, intergenic, other
Comparison to known variants	dbSNP ID, 1000 Genomes Project frequency with ethnicity, other disease-specific databases and information
Gene/coding effects	Annotated gene at the variant site, amino acid change, effect on protein sequence by the variant, algorithm scores for predicting damaging mutations

Filtering and Annotation

There are several components of annotation (Table 9.3). The confidence in the variant call is determined by sequence quality and coverage. **Coverage** is the number of times the region containing the variant is sequenced from independent fragments (read depth). Coverage is critical for confident detection of variants that are of low frequency in the sample such as somatic mutations in heterogeneous tumor tissue. Coverage of at least 500× (total of forward and reverse sequences) is recommended for detection of somatic variants.

The chromosomal and sequence location of the variant in context with the reference sequence is identified, along with the type of variant (SNP, insertion, deletion, or complex). The variant is then subjected to **filtering**. SNPs are compared to previously reported variants identified as human genome polymorphisms with the SNP rs identification number. Variants may be categorized as genetic or somatic in origin and, if genetic, as homozygous or heterozygous with the reference allele. Some variants are naturally occurring polymorphisms in particular populations. Data from the 1000 Genomes Project from major ethnic populations can be used to determine if a sequence variant is naturally present.

For gene panels and exome sequencing, variants will likely be found in gene-coding regions and adjacent intronic sequences, although intergenic areas may also be covered. The particular gene affected and the location of the variant in exon, intron, or intergenic sequences are noted. For variants found in introns, any effects of splicing are assessed. Variant effects on protein can also be estimated using algorithms such as PolyPhen and SIFT.[34] Silent variants will not change the amino acid sequence, but codon usage may have an effect on translation efficiency. Conservative amino acid substitutions or those late in the protein sequence have less effect on protein function than nonconservative mutations located early in the protein sequence. Algorithms provide scores to indicate the degree of damage to protein structure or function caused by the sequence variant. The dbNSFP database is a collection of in silico detected nonsynonymous variants.

Variants that remain after filtering may be annotated by searching in disease-specific databases, such as the Cancer Genome Atlas (TCGA), the Catalogue of Somatic Mutations in Cancer (COSMIC), My Cancer Genome, the Leiden Open (source) Variation Database (LOVD), and the Human Gene Mutation Database (HGMD). These databases and others contain population and clinical data associated with previously observed variants. The information from these databases can assist with the interpretation of the clinical effect of a variant. There are ongoing efforts to consolidate variant/disease data to ever larger and more comprehensive collections. Final reports of variants may contain information from databases, including effects on therapeutic treatments, especially targeted therapies, clinical trials, and prognosis. The clinical significance of a variant may differ with the heterogeneity of disease states as well as patient characteristics and demographics (e.g., age or gender).

Advanced Concepts

Based on professional surveys and literature reviews, a multidisciplinary group has proposed a system to categorize somatic variants in cancer.[35] It defines four tiers of variants as determined from cancer variant databases: tier I, variants with strong clinical significance; tier II, variants with potential clinical significance; tier III, variants of unknown significance (VUS); and tier IV, variants likely to be benign.

BIOINFORMATICS

Information technology has had to encompass the vast amount of data arising from the growing numbers of sequence discovery methods, especially direct sequencing and array technology. This deluge of information requires careful submission, storage, organization, and indexing of large amounts of data into databases such as those used in clinical sequencing analysis. **Bioinformatics** is the merger of biology with information technology. Part of the practice in this field is biological analysis **in silico,** that is, by computer. Bioinformatics dedicated specifically to handling sequence information is a form of **computational biology.** A list of some of the terms used in bioinformatics is shown in Table 9.4. The handling of the mountains of data being generated requires continual renewal of stored data, and a number of database programs are available for this purpose.[36,37]

Standard expression of sequence data is important for the clear communication and organized storage of sequence data. In some cases, such as in heterozygous mutations, there may be more than one base or mixed bases at the same position in the sequence. Polymorphic or heterozygous sequences are written as **consensus sequences,** or a family of sequences, with proportional representation of the polymorphic bases. The International Union of Pure and Applied Chemistry and the International Union of Biochemistry and Molecular Biology (IUB) have assigned a universal nomenclature for mixed, degenerate, or wobble bases (Table 9.5). The base designations in the IUB code are used to communicate consensus sequences and for computer input of polymorphic sequence data.

In addition to the interpretation of sequence variants, sequence information is also used in epidemiology, to speciate organisms or to find homologies within or between species. These applications involve database searches with comparisons of large regions of DNA. The **Basic Local Alignment Search Tool (BLAST)** is a system for homology searches. BLAST searches GenBank, a large database maintained by the National Center for Biotechnology Information (NCBI). Searches can be made of nucleic acid and amino acid sequences. Searches are performed by selecting a nucleotide or protein search and entering a sequence (query). Limits and parameters on the search can be added, such as the type of organisms to search (e.g., human, mouse, or other), exclusions and limits of organism or sample type, and the program. The program can optimize for highly similar sequence matches (megablast) or imperfect matches. Because sequences are directly submitted by researchers, there may be differences in the entered sequences due to the source of the sequenced material, the sequencing method, or the quality of the sequence. Selecting less-than-perfect matches will also allow cross-species matches of phylogenically conserved sequences, which can lead to the identification of important protein domains or clues to protein function.

The search will generate a number of matches or hits, with a diagram showing the alignments of the matching sequences and a color code indicating the best matches. Another section of the search results in E-values. The **E-value** (Expect value) describes the number of matches to the query by chance when searching a database of a particular size. It decreases exponentially with the quality of the match. Very low E-values (e.g., 10^{-12}) would be associated with a perfect match for a given query sequence. Further information, including the matched gene name and its organism, the source of the matched sequence and the location within that sequence, comparison of base to base or amino acid to amino acid, and plus or minus strand of the matched nucleotide sequence, are accessed by selecting the sequence or the color-coded bar in the diagram. The original submitted sequence can be accessed by selecting the sequence name.

In addition to the identification of new sequences, queries such as these are also useful for test and primer

TABLE 9.4 Bioinformatics Terminology

Term	Definition
Identity	The extent to which two sequences are the same
Alignment	Lining up two or more sequences to search for the maximal regions of identity in order to assess the extent of biological relatedness or homology
Local alignment	Alignment of some portion of two sequences
Multiple sequence alignment	Alignment of three or more sequences arranged with gaps so that common residues are aligned together
Optimal alignment	The alignment of two sequences with the best degree of identity
Conservation	Specific sequence changes (usually protein sequence) that maintain the properties of the original sequence
Similarity	The relatedness of sequences, the percent identity or conservation
Algorithm	A fixed set of commands in a computer program
Domain	A discreet portion of a protein or DNA sequence
Motif	A highly conserved short region in protein domains
Gap	A space introduced in alignment to compensate for insertions or deletions in one of the sequences being compared
Homology	Similarity attributed to descent from a common ancestor
Orthology	Homology in different species due to a common ancestral gene
Paralogy	Homology within the same species resulting from gene duplication
Query	The sequence presented for comparison with all other sequences in a selected database
Annotation	Description of functional structures, such as introns or exons in DNA or secondary structure or functional regions to protein sequences
Interface	The point of meeting between a computer and an external entity, such as an operator, a peripheral device, or a communications medium
GenBank	The genetic sequence database sponsored by the National Institutes of Health
PubMed	Search service sponsored by the National Library of Medicine that provides access to literature citations in Medline and related databases
SwissProt	Protein database sponsored by the Medical Research Council (United Kingdom)

design. Whenever a new primer or probe sequence is chosen, it is useful to query the primer or probe sequence to confirm that it belongs to the correct species and is not duplicated in multiple places in a genome. Primers and probes with multiple potential binding sites will produce mis-primes and off-target products.

Bioinformatics includes handling and updating of information for software tools and databases. The

TABLE 9.5 IUB Universal Nomenclature for Mixed Bases

Symbol	Bases	Mnemonic
A	Adenine	Adenine
C	Cytosine	Cytosine
G	Guanine	Guanine
T	Thymine	Thymine
U	Uracil	Uracil
R	A, G	puRine
Y	C, T	pYrimidine
M	A, C	aMino
K	G, T	Keto
S	C, G	Strong (3 H bonds)
W	A, T	Weak (2 H bonds)
H	A, C, T	Not G
B	C, G, T	Not A
V	A, C, G	Not T
D	A, G, T	Not C
N	A, C, G, T	aNy
X, ?	Unknown	A or C or G or T
O, -	Deletion	

accumulation of genomic and proteomic data, species and types of microorganisms based on sequences data, and variant association with disease drives the development of high-powered, reliable computer systems for storage as well as organization.

THE HUMAN GENOME PROJECT

From the first description of its double-helical structure in 1953 to the creation of the first recombinant molecule in the laboratory in 1972, DNA and the chemical nature of the arrangement of its nucleotides have attracted great interest. Gradually, this information began to accumulate, first regarding simple microorganisms and then partially in lower and higher eukaryotes. The deciphering of the human genome was a benchmark in the ongoing discovery of the molecular basis for disease and the groundwork of molecular diagnostics. In the process of solving the human DNA sequence, genomes of a variety of clinically important organisms were deciphered, advancing typing and predicting infectious disease treatment outcomes.

The first complete genome sequence of a clinically important organism was that of Epstein–Barr virus, published in 1984.[38] The 170,000-bp sequence was determined using the M13 template preparation/chain termination manual sequencing method. In 1985 and 1986, the possibility of mapping or sequencing the human genome was discussed at meetings at the University of California, Santa Cruz; Cold Spring Harbor, New York; and the Department of Energy in Santa Fe, New Mexico. The idea was controversial because of the risk that the $2 to $5 billion cost of the project might not justify the information gained, most of which would be sequences of “junk,” or non-gene-coding DNA. Furthermore, there was no available technology up to the massive task. The sequencing automation and the computer power necessary to assemble the 3 billion bases of the human genome into an organized sequence of 23 chromosomes had not yet been developed.

Nevertheless, several researchers, including Walter Gilbert (of Maxam–Gilbert sequencing), Robert Sinsheimer, Leroy Hood, David Baltimore, David Botstein, Renato Dulbecco, and Charles DeLici, saw that the project was feasible because technology was rapidly advancing toward full automation of the process. In 1982, Akiyoshi Wada had proposed automated sequencing machinery and had gotten support from Hitachi Instruments. In 1987, Smith and Hood announced the first automated DNA sequencing machine.[39] Advances in the chemistry of the sequencing procedure were accompanied by advances in the biology of DNA mapping, with methods such as pulsed-field gel electrophoresis,[40,41] restriction fragment length polymorphism analysis,[42] and transcript identification. Methods were developed to clone large (500-kbp) DNA fragments in artificial chromosomes, providing long contiguous sequencing templates.[43] Finally, application of capillary electrophoresis

TABLE 9.6 Model Organisms Sequenced During the Human Genome Project

Organism	Genome Size (Mb)	Estimated Number of Genes
Epstein–Barr virus	0.17	80
Mycoplasma genitalium	0.58	470
Haemophilus influenzae	1.8	1,740
Escherichia coli K-12	4.6	4,377
E. coli O157	5.4	5,416
Saccharomyces cerevisiae	12.5	5,770
Drosophila melanogaster	180	13,000
Caenorhabditis elegans	97	19,000
Arabidopsis thaliana	90	25,000

to DNA resolution[44–46] made the sequencing procedure even more rapid and cost-efficient.

With these developments in technology, the Human Genome Project was endorsed by the National Research Council. The National Institutes of Health (NIH) established the Office of Human Genome Research with James Watson as its head. Over the next 5 years, meetings on policy, ethics, and the cost of the project resulted in a plan to complete 20 Mb of sequence of model organisms by 2005 (Table 9.6). To organize and compare the growing amount of sequence data, the BLAST and Gene Recognition and Assembly Internet Link (GRAIL) algorithms were introduced in 1990.[47,48]

For the human sequence, the decision was made to use a composite template from multiple individuals rather than a single genome from one donor. Human DNA was donated by 100 anonymous volunteers; only 10 of these genomes were sequenced. Not even the volunteers knew if their DNA was used for the project. To ensure accurate and high-quality sequencing, all regions were sequenced 5 to 10 times.

A second project started with the same goal. In 1992, Craig Venter left the NIH to start the Institute for Genomic Research (TIGR). Venter's group completed the first sequence of a free-living organism (*Haemophilus influenzae*)[49] and the sequence of the smallest free-living organism (*Mycoplasma genitalium*).[50] Venter established a new company named Celera and proposed to complete the human genome sequence in 3 years for $300 million, faster and cheaper than the NIH project. Meanwhile, Watson had resigned as head of the NIH project and was replaced by Francis Collins. In response, the Wellcome Trust doubled its support of the NIH project. The NIH moved its completion date from 2005 to 2003, with a working draft to be completed by 2001. Thus began a competitive effort on two fronts to sequence the human genome.

The two projects approached the sequencing differently (Fig. 9.25). The NIH method (hierarchical shotgun sequencing) was to start with sequences of known regions in the genome and "walk" further away into the chromosomes, always aware of where the newly generated sequences belonged in the human genome map. Venter and the researchers working with Celera—Gene Meyers, Jane Rogers, Robert Millman, John Sulston, and Todd Taylor—had a different idea. Their approach *(whole-genome shotgun sequencing)* was to start with 10 equivalents of the human genome cut into small fragments and randomly sequence the lot. Then, powerful computers would find overlapping sequences and use those to assemble the billions of bases of sequence into their proper chromosomal locations.

Initially, the Celera approach was met with skepticism. The human genome contains large amounts of repeated sequences, some of which are very difficult to sequence and even more difficult to map properly. A random sequencing method would repeatedly cover areas of the genome that are more easily sequenced and miss more difficult regions. Moreover, assembly of the whole sequence from scratch with no chromosomal landmarks would take a prohibitive amount of computer power. Nonetheless, Celera began to make headway (some alleged with the help of the publicly published sequences from the NIH), and eventually, the NIH project modified its approach to include both methods.

Over the next months, some efforts were made toward combining the two projects, but these efforts broke down over disagreements over database policy and release of completed sequences. The result of the competition was that the rough draft of the sequence was completed by

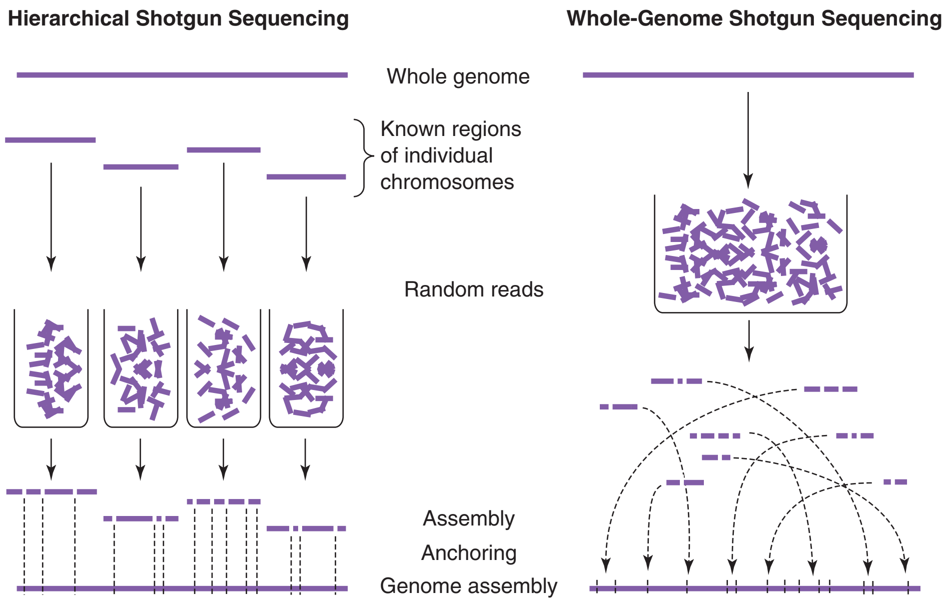

Hierarchical Shotgun Sequencing
Whole-Genome Shotgun Sequencing
Whole genome
Known regions of individual chromosomes
Random reads
Assembly
Anchoring
Genome assembly

FIGURE 9.25 Comparison of two approaches for sequencing of the human genome. The hierarchical shotgun approach taken by the NIH *(left)* was to sequence from known regions so that new sequences could easily be located in the genome. The Celera whole-genome shotgun approach *(right)* was to sequence random fragments from the entire genome and then assemble the complete sequence with computers.

both projects earlier than either group had proposed, in June 2000. A joint announcement was made, and both groups published their versions of the genome, the NIH version in the journal *Nature*[51] and the Celera version in the journal *Science*.[52]

The sequence completed in 2000 was a rough draft of the genome; that is, there were still areas of missing sequence and sequences yet to be placed. Only chromosomes 21 and 22, the smallest of the chromosomes, had been fully completed. In the ensuing years, the finished sequences of each chromosome have been released (Table 9.7).

Remaining errors, gaps, and complex gene rearrangements will take years to resolve.[53] Detailed analysis of an individual genome will require sequencing of both homologs of each chromosome.[54] Even with the rough draft, interesting characteristics of the human genome were revealed. The size of the entire genome is 2.91 Gbp (2.91 billion bp). The genome was initially calculated as 54% AT and 38% GC, with 8% of the bases still to be determined. Chromosome 2 is the most GC-rich chromosome (66%), and chromosome X has the fewest GC base pairs (25%). A most surprising discovery was that the number of genes, estimated to be from 20,000 to 30,000, was much lower than expected. The average size of a human gene is 27 kbp. Chromosome 19 is the most gene-rich per unit length (23 genes/Mbp). Chromosomes 13 and Y have the fewest genes per base pair (5 genes/Mbp). Only about 2% of the sequences code for genes. Between 30% and 40% of the genome consists of repeat sequences. There is one single-base difference between two random individuals found approximately every 1,000 bases along the human DNA sequence. More detailed information, databases, references, and updated information are available at http://www.ncbi.nlm.nih.gov/.

The promise of the Human Genome Project for molecular diagnostics can be appreciated with the example of the discovery of the gene involved in cystic fibrosis. Seven years of work were required for discovery of this gene. With proper mapping information, a gene for any disease can now be found by computer, already sequenced, in a matter of minutes. Of course, all genetic diseases are not due to the malfunction of

TABLE 9.7 Completed Chromosomes

Chromosome	Completion Date
21	December 1999
22	May 2000
20	December 2001
14	January 2003
Y	June 2003
7	July 2003
6	October 2003
13	March 2004
19	March 2004
9	May 2004
9	May 2004
5	September 2004
16	December 2004
18	January 2005
X	March 2005
2	April 2005
4	April 2005
8	January 2006
11	March 2006
15	March 2006
12	March 2006
17	April 2006
3	April 2006
1	May 2006

a single gene. In fact, most diseases and normal states are driven by a combination of genes as well as by environmental influences. Without the rich information afforded by the sequence of the human genome, identification of these multicomponent diseases would be almost impossible. Ten years after the announcement of the completion of the Human Genome Project, almost 200 human genomes had been sequenced. Although the information gathered from the sequencing effort had not yielded the benefits to human health expected at the start of the project, it increased the appreciation of the vast complexity of genes and their regulation.[55]

Variant Associations With Phenotype

The Human Haplotype Mapping Project

As the Human Genome Project moved toward completion, another project was launched to further define the relationship between gene sequence and disease. This was the Human Haplotype Mapping, or HapMap, Project.[56] The goal of the project was to find blocks of sequences that are inherited together, marking particular traits and possibly disease-associated genetic lesions. The haplotype approach would reduce the number of polymorphisms required to examine the entire collection of genome/phenotype associations from the 10 million polymorphisms that exist to roughly 500,000 haplotypes. The HapMap Project revealed more than 1,000 disease-associated regions of the genome, covering commonly occurring conditions such as coronary artery disease and diabetes. With advances in technology, and the ability to generate sequence information, however, HapMap data has mostly been supplanted by higher-throughput data-gathering methods. As a result, the NCBI has planned to retire the HapMap project site because other resources, such as the 1000 Genomes Project, have become more comprehensive references for population genomics.

The 1000 Genomes Project

The 1000 Genomes Project provides a resource of structural variants in different populations.[57] The project has reconstructed the genomes of over 2,504 individuals from 26 populations by whole-genome sequencing, deep exome sequencing, and dense microarray genotyping in laboratories in the United States, United Kingdom, China, and Germany. Over 88 million variants (84.7 million SNPs, 3.6 million short insertions/deletions, and 60,000 structural variants) were verified. The resulting database includes more than 99% of single-nucleotide variants with a frequency of greater than 1%. Data from the 1000 Genomes Project is a component of NGS variant assessment, providing more patient-specific interpretation of the clinical significance of variants. All variants from the 1000 Genomes Project are submitted to archives such as dbSNP.

The majority of HapMap SNPs are found in the 1000 Genomes Project.[58] Sites from HapMap that aren't found by the 1000 Genomes Project may be false discoveries by HapMap, the latter being based on microarray technology. Thus, there are a lot of SNPs from NGS projects that are not reported in HapMap.

The technology developed as part of the Human Genome Project made sequencing a routine method in the clinical laboratory. Small, cost-effective sequencers are available for rapid sequencing. In the clinical laboratory, sequencing is actually resequencing, or repeated analysis of the same sequence region, to detect mutations or to type microorganisms, making the task even more routine. The technology continued to develop, reducing the cost and labor of sequencing to detect multicomponent diseases or to predict predisposition to disease. Massive parallel or next-generation sequencing has supplemented and/or replaced Sanger sequencing in many clinical laboratories, and even this technology has evolved into lower-cost, user-friendly protocols. Accurate and comprehensive sequence analysis is one of the most promising areas of molecular diagnostics.

STUDY QUESTIONS

1. Read 5′ to 3′ the first 15 bases of the sequence in the gel on the right in Figure 9.7 (p. 229).

2. After an automated dye primer sequencing run, the electropherogram displays consecutive peaks of the following colors:
   red, red, black, green, green, blue, black, red, green, black, blue, blue, blue
   If the computer software displays the fluors from ddATP as green, ddCTP as blue, ddGTP as black, and ddTTP as red, what is the sequence of the region given?

3. A dideoxy sequencing electropherogram displays bright (high, wide) peaks of fluorescence, obliterating some of the sequencing peaks. What is the most likely cause of this observation? How might it be corrected?

4. In a manual sequencing reaction, the sequencing ladder on the polyacrylamide gel is very bright and readable at the bottom of the gel, but the larger (slower-migrating) fragments higher up are very faint. What is the most likely cause of this observation? How might it be corrected?

5. In an analysis of the *TP53* gene for mutations, the following sequences were produced. For each sequence, write the expected sequence of the opposite strand that would confirm the presence of the mutations detected.

   5′TATCTGTTCACTTGTGCCCT3′ (Normal)
   5′TATCTGTTCATTTGTGCCCT3′ (Homozygous substitution)
   5′TATCTGT(T/G)CACTTGTGCCCT3′ (Heterozygous substitution)
   5′TATCTGTT(C/A)(A/C)(C/T)T(T/G)(G/T)(T/G) (G/C)CC(C/T) . . . 3′ (Heterozygous deletion)

6. A sequence, TTGCTGCGCTAAA, may be methylated at one or more of the cytosine residues. After bisulfite sequencing, the following results are obtained:

   Bisulfite treated: TTGUTGCGUTAAA
   Write the sequence showing the methylated cytosines as C^{Me}.

7. In a pyrosequencing readout, the graph shows peaks of luminescence corresponding to the addition of the following nucleotides:

   dT peak, dC peak (double height), dT peak, dA peak
   What is the sequence?

8. Why is it necessary to add adenosine residues in vitro to ribosomal RNA before capture for sequencing?

9. Which of the following is next-generation sequencing?
    - a. Maxam–Gilbert
    - b. Tiled microarray
    - c. Dideoxynucleotide chain terminator sequencing
    - d. Reversible dye terminator sequencing

10. Which of the following projects would require next-generation sequencing?
    - a. Mapping a mutation in the hemochromatosis gene
    - b. Sequencing a viral genome
    - c. Characterizing a diverse microbial population
    - d. Typing a single bacterial colony

References

1. Sheikine Y, Kuo FC, Lindeman NI. Clinical and technical aspects of genomic diagnostics for precision oncology. *Journal of Clinical Oncology* 2017;35:929–933.
2. Maxam A, Gilbert W. Sequencing end-labeled DNA with base-specific chemical cleavage. *Methods in Enzymology* 1980;65: 499–560.
3. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain terminating inhibitors. *Proceedings of the National Academy of Sciences* 1977;74:5463–5467.
4. Metzker ML, Lu J, Gibbs RA. Electrophoretically uniform fluorescent dyes for automated DNA sequencing. *Science* 1996;271: 1420–1422.
5. Lewis E, Haaland WC, Nguyen F, Heller DA, Allen MJ, MacGregor RR, Berger CS, Willingham B, Burns LA, Scott GB, Kittrell C, Johnson BR, Curl RF, Metzker ML. Color-blind fluorescence detection for four-color DNA sequencing. *Proceedings of the National Academy of Sciences* 2005;102:5346–5351.
6. Ronaghi M, Uhlen M, Nyren P. A sequencing method based on real-time pyrophosphate. *Science* 1998;281:363–365.
7. Harrington CT, Lin E, Olson MT, Eshleman JR. Fundamentals of pyrosequencing. *Archives of Pathology & Laboratory Medicine* 2013;137:1296–1303.
8. Insuasti-Beltran G, Gale JM, Wilson CS, Foucar K, Czuchlewski DR. Significance of MYD88 L265P mutation status in the subclassification of low-grade B-cell lymphoma/leukemia. *Archives of Pathology & Laboratory Medicine* 2015;139:1035–1041.
9. Lin S, Desmond EP. Molecular diagnosis of tuberculosis and drug resistance. *Clinical Laboratory Medicine* 2014;34:297–314.
10. Stürmer M, Reinheimer C. Description of two commercially available assays for genotyping of HIV-1. *Intervirology* 2012;55: 134–137.
11. Marsh S. Pyrosequencing applications. *Methods in Molecular Biology* 2007;373:15–24.
12. Shiraishi M, Hayatsu H. High-speed conversion of cytosine to uracil in bisulfite genomic sequencing analysis of DNA methylation. *DNA Research* 2004;11:409–415.
13. Weller M, Tabatabai G, Kästner B, Felsberg J, Steinbach JP, Wick A, Schnell O, Hau P, Herrlinger U, Sabel MC, Wirsching HG, Ketter R, Bähr O, Platten M, Tonn JC, Schlegel U, Marosi C, Goldbrunner R, Stupp R, Homicsko K, Pichler J, Nikkhah G, Meixensberger J, Vajkoczy P, Kollias S, Hüsing J, Reifenberger G, Wick W; DIRECTOR Study Group. MGMT promoter methylation is a strong prognostic biomarker for benefit from dose-intensified temozolomide rechallenge in progressive glioblastoma: the DIRECTOR trial. *Clinical Cancer Research* 2015;21:2057–2064.
14. Weller M, Stupp R, Reifenberger G, Brandes AA, van den Bent MJ, Wick W, Hegi ME. MGMT promoter methylation in malignant gliomas: ready for personalized medicine? *Nature Reviews Neurology* 2010;6:39–51.
15. Cocquet J, Chong A, Zhang G, Veitia RA. Reverse transcriptase template switching and false alternative transcripts. *Genomics* 2006;88:127–131.
16. Roberts J, Preston BD, Johnston LA, Soni A, Loeb LA, Kunkel TA. Fidelity of two retroviral reverse transcriptases during DNA-dependent DNA synthesis in vitro. *Molecular and Cellular Biology* 1989;9:469–476.
17. Braslavsky I, Hebert B, Kartalov E, Quake SR. Sequence information can be obtained from single DNA molecules. *Proceedings of the National Academy of Sciences* 2003;100:3960–3964.
18. Ozsolak F, Platt AR, Jones DR, Reifenberger JG, Sass LE, McInerney P, Thompson JF, Bowers J, Jarosz M, Milos PM. Direct RNA sequencing. *Nature* 2009;461:814–818.
19. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics* 2016;17:333–351.
20. Mardis E. The impact of next-generation sequencing technology on genetics. *Trends in Genetics* 2008;24:133–141.
21. Pop M, Salzberg SL. Bioinformatics challenges of new sequencing technology. *Trends in Genetics* 2008;24:142–149.
22. Fuller C, Middendorf L, Benner SA, Church GM, Harris T, Huang X, Jovanovich SB, Nelson JR, Schloss JA, Schwartz DC, Vezenov DV. The challenges of sequencing by synthesis. *Nature Biotechnology* 2009;27:1013–1023.
23. Jennings LJ, Arcila ME, Corless C, Kamel-Reid S, Lubin IM, Pfeifer J, Temple-Smolkin RL, Voelkerding KV, Nikiforova MN. Guidelines for validation of next-generation sequencing based oncology panels. A joint consensus recommendation of the Association for Molecular Pathology and College of American Pathologists. *Journal of Molecular Diagnostics* 2017;19(3):341–365.
24. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M, Hoon J, Simons JF, Marran D, Myers JW, Davidson JF, Branting A, Nobile JR, Puc BP, Light D, Clark TA, Huber M, Branciforte JT, Stoner IB, Cawley SE, Lyons M, Fu Y, Homer N, Sedova M, Miao X, Reed B, Sabina J, Feierstein E, Schorn M, Alanjary M, Dimalanta E, Dressman D, Kasinskas R, Sokolsky T, Fidanza JA, Namsaraev E, McKernan KJ, Williams A, Roth GT, Bustillo J. An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 2011;475:348–352.

25. Ruparel H, Bi L, Li Z, Bai X, Kim DH, Turro NJ, Ju J. Design and synthesis of a 3′-O-allyl photocleavable fluorescent nucleotide as a reversible terminator for DNA sequencing by synthesis. *Proceedings of the National Academy of Sciences* 2005;102: 5932–5937.

26. Xue Y, Ankala A, Wilcox WR, Hegde MR. Solving the molecular diagnostic testing conundrum for Mendelian disorders in the era of next-generation sequencing: single-gene, gene panel, or exome/genome sequencing. *Genetics in Medicine* 2015;17:444–451.

27. Jacob H, Abrams K, Bick DP, Brodie K, Dimmock DP, Farrell M, Geurts J, Harris J, Helbling D, Joers BJ, Kliegman R, Kowalski G, Lazar J, Margolis DA, North P, Northup J, Roquemore-Goins A, Scharer G, Shimoyama M, Strong K, Taylor B, Tsaih SW, Tschannen MR, Veith RL, Wendt-Andrae J, Wilk B, Worthey EA. Genomics in clinical practice: lessons from the front lines. *Science and Translational Medicine* 2013;5:1–5.

28. Manolio T, Fowler DM, Starita LM, Haendel MA, MacArthur DG, Biesecker LG, Worthey E, Chisholm RL, Green ED, Jacob HJ, McLeod HL, Roden D, Rodriguez LL, Williams MS, Cooper GM, Cox NJ, Herman GE, Kingsmore S, Lo C, Lutz C, MacRae CA, Nussbaum RL, Ordovas JM, Ramos EM, Robinson PN, Rubinstein WS, Seidman C, Stranger BE, Wang H, Westerfield M, Bult C. Bedside back to bench: building bridges between basic and clinical genomic research. *Cell* 2017;169:6–12.

29. Samorodnitsky E, Jewell BM, Hagopian R, Miya J, Wing MR, Lyon E, Damodaran S, Bhatt D, Reeser JW, Datta J, Roychowdhury S. Evaluation of hybridization capture versus amplicon-based methods for whole-exome sequencing. *Human Mutation* 2015;36: 903–914.

30. Misyura M, Zhang T, Sukhai MA, Thomas M, Garg S, Kamel-Reid S, Stockley TL. Comparison of next-generation sequencing panels and platforms for detection and verification of somatic tumor variants for clinical diagnostics. *Journal of Molecular Diagnostics* 2016;18:842–850.

31. Harismendy O, Ng PC, Strausberg RL, Wang X, Stockwell TB, Beeson KY, Schork NJ, Murray SS, Topol EJ, Levy S, Frazer KA. Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biology* 2009;10:R32.

32. Lu H, Giordano F, Ning Z. Oxford nanopore MinION sequencing and genome assembly. *Bioinformatics* 2016;14:265–279.

33. Karthikeyan S, Bawa PS, Srinivasan S. hg19K: addressing a significant lacuna in hg19-based variant calling. *Molecular Genetics and Genomic Medicine* 2016;5:15–20.

34. Flanagan S, Patch AM, Ellard S. Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. *Genet Test Mol Biomarkers* 2010;14(4):533–537.

35. Li M, Datto M, Duncavage EJ, Kulkarni S, Lindeman NI, Roy S, Tsimberidou AM, Vnencak-Jones CL, Wolff DJ, Younes A, Nikiforova MN. Standards and guidelines for the interpretation and reporting of sequence variants in cancer: a joint consensus recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. *Journal of Molecular Diagnostics* 2017;9:4–23.

36. Chojnacki S, Cowley A, Lee J, Foix A, Lopez R. Programmatic access to bioinformatics tools from EMBL-EBI update: 2017. *Nucleic Acids Research* 2017;45(W1):W550–W553.

37. Niu S, Yang J, McDermaid A, Zhao J, Kang Y, Ma Q. Bioinformatics tools for quantitative and functional metagenome and metatranscriptome data analysis in microbes. *Briefings in Bioinformatics* 2017. doi:10.1093/bib/bbx051

38. Baer R, Bankier AT, Biggin MD, Deininger PL, Farrell PJ, Gibson TJ, Hatfull G, Hudson GS, Satchwell SC, Seguin C. DNA sequence and expression of the B95-8 Epstein-Barr virus genome. *Nature* 1984;310:207–211.

39. Hood L, Hunkapiller MW, Smith LM. Automated DNA sequencing and analysis of the human genome. *Genomics* 1987;1:201–212.

40. Schwartz D, Cantor CR. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell* 1984;37: 67–75.

41. Van der Ploeg L, Schwartz DC, Cantor CR, Borst P. Antigenic variation in Trypanosoma brucei analyzed by electrophoretic separation of chromosome-sized DNA molecules. *Cell* 1984;37:77–84.

42. Donis-Keller H, Green P, Helms C, Cartinhour S, Weiffenbach B, Stephens K, Keith TP, Bowden DW, Smith DR, Lander ES. A genetic linkage map of the human genome. *Cell* 1987;51:319–337.

43. Riethman H, Moyzis RK, Meyne J, Burke DT, Olson MV. Cloning human telomeric DNA fragments into *Saccharomyces cerevisiae* using a yeast-artificial-chromosome vector. *Proceedings of the National Academy of Sciences* 1989;86:6240–6244.

44. Luckey J, Drossman H, Kostichka AJ, Mead DA, D'Cunha J, Norris TB, Smith LM. High speed DNA sequencing by capillary electrophoresis. *Nucleic Acids Research* 1990;18:4417–4421.

45. Karger A. Separation of DNA sequencing fragments using an automated capillary electrophoresis instrument. *Electrophoresis* 1996;17:144–151.

46. Chen D, Swerdlow HP, Harke HR, Zhang JZ, Dovichi NJ. Low-cost, high-sensitivity laser-induced fluorescence detection for DNA sequencing by capillary gel electrophoresis. *Journal of Chromatography* 1991;559:237–246.

47. Altschul S, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *Journal of Molecular Biology* 1990;215: 403–410.

48. Xu Y, Mural RJ, Uberbacher EC. Constructing gene models from accurately predicted exons: an application of dynamic programming. *Computer Applications in the Biosciences* 1994;10:613–623.

49. Fleischmann R, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb J-F, Dougherty BA, Merrick JM, McKenney K, Sutton GG, FitzHugh W, Fields CA, Gocayne JD, Scott JD, Shirley R, Liu L-I, Glodek A, Kelley JM, Weidman JF, Phillips CA, Spriggs T, Hedblom E, Cotton MD, Utterback TR, Hanna MC, Nguyen DT, Saudek DM, Brandon RC, Fine LD, Fritchman JL, Fuhrmann JL, Geoghagen NSM, Gnehm CL, McDonald LA, Small KV, Fraser CM, Smith HO, Venter JC. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 1995;269:496–512.

50. Fraser C, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, Bult CJ, Kerlavage AR, Sutton G, Kelley JM, Fritchman RD, Weidman JF, Small KV, Sandusky M, Fuhrmann J, Nguyen D, Utterback TR, Saudek DM, Phillips CA, Merrick JM, Tomb JF, Dougherty BA, Bott KF, Hu PC, Lucier TS, Peterson SN, Smith HO, Hutchison CA, Venter JC. The minimal gene complement of *Mycoplasma genitalium*. *Science* 1995;270:397–403.

51. Lander E, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ, Szustakowki J. Initial sequencing and analysis of the human genome. *Nature* 2001;409:860–921.

52. Venter J, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigó R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X. The sequence of the human genome. *Science* 2001;291:1304–1351.

53. Dolgin E. The genome finishers. *Nature* 2009;462:843–845.

54. Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G, Lin Y, MacDonald JR, Pang AW, Shago M, Stockwell TB, Tsiamouri A, Bafna V, Bansal V, Kravitz SA, Busam DA, Beeson KY, McIntosh TC, Remington KA, Abril JF, Gill J, Borman J, Rogers YH, Frazier ME, Scherer SW, Strausberg RL, Venter JC. The diploid genome sequence of an individual human. *PLoS Biology* 2007;5:2113–2145.

55. Hayden E. Life is complicated. *Nature* 2010;464:664–667.

56. Consortium IH. The International HapMap Project. *Nature* 2003;426:789–796.

57. Consortium TGP. A global reference for human genetic variation. *Nature* 2015;526(7571):68–74.

58. Buchanan C, Torstenson ES, Bush WS, Ritchie MD. A comparison of cataloged variation between International HapMap Consortium and 1000 Genomes Project data. *Journal of the American Medical Association* 2012;19:289–294.

Section III

Techniques in the Clinical Laboratory

Chapter 10

DNA Polymorphisms and Human Identification

Outline

TYPES OF POLYMORPHISMS
RFLP TYPING
- *Genetic Mapping With RFLPs*
- *RFLP and Parentage Testing*
- *Human Identification Using RFLPs*
    - DNA Fingerprinting With RFLP

STR TYPING BY PCR
- *STR Analysis*
    - STR Nomenclature
    - Gender Identification
    - Analysis of Test Results
    - Genotyping
    - Matching of Profiles
    - Allelic Frequencies in Paternity Testing
    - Sibling Tests

Y-STR
- *Matching With Y-STRs*

LINKAGE ANALYSIS
BONE MARROW ENGRAFTMENT TESTING USING DNA POLYMORPHISMS
- *PSTR Testing*
- *Post-Transplant Engraftment Testing*

QUALITY ASSURANCE OF TISSUE SECTIONS USING STR
SINGLE-NUCLEOTIDE POLYMORPHISMS
- *The Human Haplotype Mapping (HapMap) Project*

MITOCHONDRIAL DNA POLYMORPHISMS
OTHER IDENTIFICATION METHODS
- *Protein-Based Identification*
- *Epigenetic Profiles*

Objectives

10.1 Compare and contrast different types of polymorphisms.

10.2 Define *restriction fragment length polymorphism* (RFLP), and discuss how RFLPs are used in genetic mapping, parentage testing, and human identification.

10.3 Describe short tandem repeat (STR) structure and nomenclature.

10.4 Describe the use of STR in parentage testing and the amelogenin locus for gender identification.

10.5 Explain matching probabilities and the contribution of allele frequencies to the certainty of matching.

10.6 Describe the use of Y-STR in forensic and lineage studies.

10.7 Use STR for linkage analysis.
10.8 Give examples of the use of STR for bone marrow engraftment monitoring.
10.9 Show how STR may be used for quality assurance of histological sections.
10.10 Define *single-nucleotide polymorphism* (SNP), and explain the potential use of SNPs in disease gene mapping.
10.11 Discuss mitochondrial DNA typing.
10.12 Identify a protein profile from a reference database.
10.13 Predict the effect of aging on epigenetic (DNA methylation) profiles.

Polymorphisms are variations of DNA sequences that are shared by a certain percentage of a population. These sequences range from a single base pair to thousands of base pairs.

TYPES OF POLYMORPHISMS

The probability of polymorphic DNA in humans is great due to the relatively large size of the human genome, 98% of which does not code for genes. At the nucleotide-sequence level, it is estimated that genome sequences differ by at least one nucleotide every 1,000 to 1,500 bases. These single-nucleotide differences, or **single-nucleotide polymorphisms (SNPs),** may occur in gene-coding regions or in intergenic sequences.

Polymorphisms are more frequent in some areas of the genome than in others. The human leukocyte antigen (HLA) locus is a familiar example of a highly polymorphic region of human DNA. The variable nucleotide sequences in this locus code for peptides that establish self-identity of the immune system. The extent of similarity or compatibility between the immune systems of transplant recipients and potential donors can thus be determined by comparing DNA sequences. Some human sequence polymorphisms affect many base pairs. Large blocks of repeated sequences may be inverted, deleted, or duplicated from one individual to another. **Long interspersed nucleotide elements (LINEs)** are highly repeated sequences, 6 to 8 kbp in length, that contain RNA polymerase promoters and open reading frames related to the reverse transcriptase of retroviruses. There are more than 500,000 of these LINE-1 (L1) elements, making up more than 15% of the human genome. There are even more **short interspersed nucleotide elements (SINEs)** scattered over the genome. SINEs, 0.3 kbp in size, are present in over 1,000,000 copies per genome. SINEs include **Alu elements,** named for harboring recognition sites for the *Alu*I restriction enzyme. There are well in excess of 1 million Alu elements, accounting for almost 11% of the human genome.[1] The majority of transcribed genes contain Alu elements in their introns. Alu elements have cryptic splice and polyadenylation sites, which can become activated through the accumulation of mutations and lead to alternative splicing of RNAs or premature termination of translation. LINEs and SINEs are also known as mobile elements or transposable elements. They are copied and spread by recombination and reverse transcription and may be responsible for the formation of **pseudogenes** (intronless, nonfunctional copies of active genes) throughout the human genome.

Shorter blocks of repeated sequences also undergo expansion or shrinkage through generations. Examples of the latter are **short tandem repeats (STRs)** and **variable-number tandem repeats (VNTRs).**

SNPs, larger sequence variants, and tandem repeats can be detected by observing changes in the restriction map of a DNA region. Analysis of restriction fragments by Southern blot reveals **restriction fragment length polymorphisms (RFLPs).** Particular types of polymorphisms, specifically SNPs, VNTRs, STRs, and RFLPs, are routinely used in the laboratory (Table 10.1).

Historical Highlights

In the 1920s, scientists realized that blood type (A, B, AB, or O) is inherited and could be used for parentage testing. This limited testing could only exclude a falsely alleged father. But soon after, the use of other proteins on the surface of the red blood cell (Rh, Kell, and Duffy blood group systems) was introduced. The power of these serological tests was only marginally better than that of the ABO system. Forty years later, the polymorphic HLAs were implemented for parentage and identity testing, coupled with the ABO and serological testing.

TABLE 10.1 Types of Useful Polymorphisms and Laboratory Methods

Polymorphism	Structure	Detection Method
RFLP	One or more nucleotide changes that affect the size of restriction enzyme products	Southern blot
VNTR	Repeats of 10–50 base sequences in tandem	Southern blot, PCR
STR	Repeats of 1–10 base sequences in tandem	PCR
SNP	Alterations of a single nucleotide	Sequencing, other

RFLP TYPING

RFLPs were the original DNA targets used for gene mapping, human identification, and parentage testing. The first polymorphic RFLP was described in 1980. RFLPs are observed as differences in the sizes and number of fragments generated by restriction enzyme digestion of DNA (Fig. 10.1). Fragment sizes vary as a result of changes in the nucleotide sequence in or between the recognition sites of a restriction enzyme. Nucleotide changes may also destroy, change, or create restriction enzyme sites, altering the number of fragments.

The first step in using RFLPs is to construct a restriction enzyme map of the DNA region under investigation. Once the restriction map is known, the number and sizes of the restriction fragments of a test DNA region cut with restriction enzymes are compared with the number and sizes of fragments expected based on the restriction map. Polymorphisms are detected by observing fragment numbers and sizes different from those expected from the reference restriction map. An example of a polymorphism in a restriction site is shown in Figure 10.2. In a theoretical linear piece of DNA, loss of the recognition site for the enzyme (*Bgl*II in the figure) results in alteration of the size and number of bands detected after gel electrophoresis.

RFLP typing in humans required the use of the Southern blot technique. DNA was cut with restriction enzymes, resolved by gel electrophoresis, and blotted to a membrane. Probes to specific regions of DNA containing potential RFLPs were then hybridized to the DNA

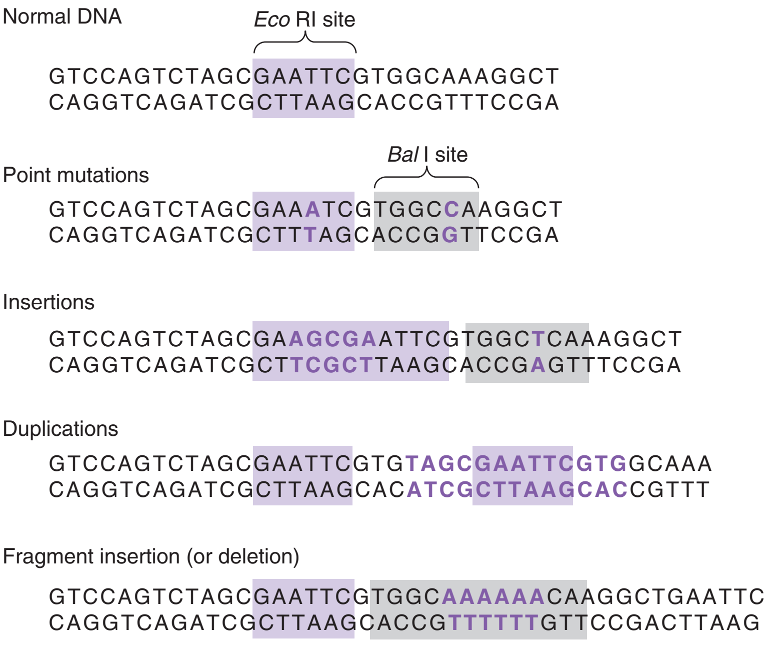
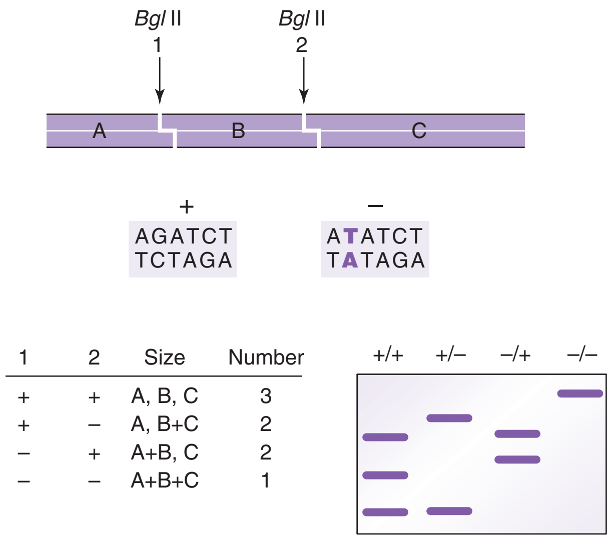
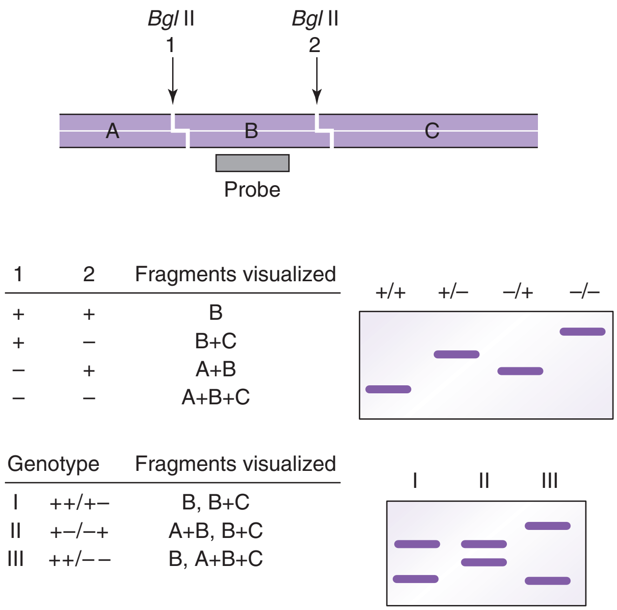

Normal DNA
Eco RI site
GTCCAGTCTAGCGAATTCGTGGCAAAGGCT
CAGGTCAGATCGCTTAAGCACCGTTTCCGA
Point mutations
Bal I site
GTCCAGTCTAGCGAAATCGTGGCCAAGGCT
CAGGTCAGATCGCTTTAGCACCGGTTCCGA
Insertions
GTCCAGTCTAGCGAAGCGAATTCGTGGCTCAAAGGCT
CAGGTCAGATCGCTTCGCTTAAGCACCGAGTTTCCGA
Duplications
GTCCAGTCTAGCGAATTCGTGTAGCGAATTCGTGGCAAA
CAGGTCAGATCGCTTAAGCACATCGCTTAAGCACCGTTT
Fragment insertion (or deletion)
GTCCAGTCTAGCGAATTCGTGGCAAAAAACAAGGCTGAATTC
CAGGTCAGATCGCTTAAGCACCGTTTTTTGTTCCGACTTAAG

FIGURE 10.1 Types of DNA sequence alterations that change restriction fragment lengths. The normal sequence *(top)* has an *Eco*R1 site (GAATTC). Single-base changes (point mutations, second line) can destroy the *Eco*R1 site or create a new restriction site, as can insertions, duplications, or deletions of any number of bases *(third through fifth lines)*. Insertions, duplications, and deletions between two restriction sites change fragment size without affecting the restriction sites themselves.



1	2	Size	Number
+	+	A, B, C	3
+	–	A, B+C	2
–	+	A+B, C	2
–	–	A+B+C	1

FIGURE 10.2 A linear piece of DNA with two polymorphic *Bgl*II restriction enzyme sites, designated as 1 and 2, will yield different fragment sizes, depending on the presence of neither, either, or both of the restriction sites. For instance, a G to T mutation will change the sequence of the normal site (+) to one not recognized by the enzyme (–). The presence or absence of the polymorphic sites is evident from the number and size of the fragments after cutting the DNA with *Bgl*II (bottom right).



1	2	Fragments visualized
+	+	B
+	–	B+C
–	+	A+B
–	–	A+B+C

Genotype		Fragments visualized
I	++/+–	B, B+C
II	+–/–+	A+B, B+C
III	++/––	B, A+B+C

FIGURE 10.3 Using a Southern blot to probe for RFLP. With the same region shown in Figure 10.2, only the fragments with complementary sequences to a probe to the B region *(top)* can be visualized. The bottom panel shows a diploid genotype where homologous chromosomes carry different RFLP alleles.

on the membrane to determine the size of the resulting bands. Figure 10.3 shows the pattern of bands resulting from a Southern blot analysis of the RFLP in the linear fragment from Figure 10.2. Even if the probe does not detect all of the restriction fragments, the polymorphisms can still be identified.

DNA is inherited as one haploid chromosome complement from each parent. Each chromosome carries its polymorphisms so that the offspring inherits a combination of the parental polymorphisms. When visualized as fragments that hybridize to a probe of a polymorphic region, the band patterns represent the combination of RFLPs inherited from each parent. Due to recombination and random assortment, each person has a unique set of RFLPs, half inherited maternally and half paternally. Every genotype will yield a descriptive band pattern, as shown in Figure 10.3.

Over many generations, mutations, intra- and interchromosomal recombination, gene conversion, and other genetic events have increased the diversity of DNA sequences. One consequence of this genetic diversity is that a single **locus,** that is, a gene or region of DNA, will have several versions, or **alleles.** Human beings are diploid with two copies of every locus. In other words, each person has two alleles of each locus. If these alleles are the same, the locus is **homozygous;** if the two alleles are different, the locus is **heterozygous.**

Depending on the extent of diversity or polymorphism of a locus, any two people can share the same alleles or have different alleles. More closely related individuals are likely to share more alleles than unrelated persons. In the examples shown in Figure 10.3, (+ +), (+ –), (– +), and (– –) describe the presence (+) or absence (–) of *Bgl*II sites making up four alleles of the locus detectable by Southern blot. In the illustration, genotypes I and II both have the (+ –) allele on one chromosome, but genotype I has (+ +), and genotype II has (– +) on the other chromosome. This appears in the Southern blot results as one band of equal size between the two genotypes and one band that is a different size.

Two individuals can share both alleles at a single locus, but the chances of two individuals, except for identical twins, sharing the same alleles decrease 10-fold with each additional locus tested.[2]

More than 2,000 RFLP loci have been described in human DNA. The uniqueness of the collection of polymorphisms in each individual is the basis for human identification at the DNA level. Detection of RFLP by Southern blot made positive paternity testing and human identification possible for the first time.

To optimize the discriminatory capacity of RFLP testing, restriction enzymes that cut human DNA frequently were used for RFLP tests. RFLP protocols for human identification in most North American laboratories used the restriction enzyme *Hae*III for fragmentation of genomic DNA. Many European laboratories used the *Hinf*I enzyme. All of these enzymes cut DNA frequently enough to reveal polymorphisms in multiple locations throughout the genome. To regulate results from independent laboratories, the Standard Reference Material (SRM) DNA Profiling Standard for RFLP analysis was released in 1992. The SRM supplies cell pellets, genomic DNA, gel standards, precut DNA, electrophoresis materials, molecular-weight markers, and certified values for final analysis. These materials, currently provided by the National Institute of Standards and Technology (NIST), were designed to maintain the reproducibility of the RFLP process across laboratories.

Genetic Mapping With RFLPs

Polymorphisms are inherited in a Mendelian fashion, and the locations of many polymorphisms in the genome are known. Therefore, polymorphisms can be used as landmarks, or markers, in the genome to determine the location of other genes. In addition to showing clear family history or direct identification of a genetic factor, one can confirm that a disease has a genetic component by demonstrating a close genetic association or linkage to a known marker. Formal statistical methods are used to determine the probability that an unknown gene is located close to a known marker in the genome. The more frequently a particular polymorphism is present in persons with a disease phenotype, the more likely an affected gene is located close to the polymorphism. This is the basis for linkage mapping and one of the ways genetic components of disease are identified.

Historical Highlights

Mary Claire King used RFLP to map one of the genes mutated in inherited breast cancer.[3,4] Following extended families with high incidence of breast and ovarian cancer, she found particular RFLP always present in affected family members. Because the location in the genome of the RFLP was known (17q21), the *BRCA1* gene was thereby mapped to this position on the long arm of chromosome 17.

RFLP and Parentage Testing

In diploid organisms, chromosomal content is inherited, half from each parent. This includes the DNA polymorphisms located throughout the genome. Taking advantage of the unique combination of RFLP in each individual, one can infer a parent's contribution of alleles to a child from the combination of alleles in the child and those of the other parent. The fragment sizes of an individual are a combination of those from each parent, as illustrated in Figure 10.4. In a paternity test, the alleles or fragment

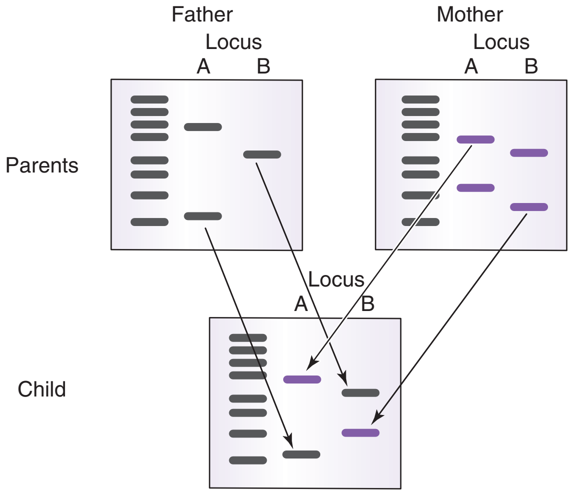


FIGURE 10.4 RFLP inheritance. Two different genetic regions, or loci, are shown, locus A and locus B. There are several versions or alleles of each locus. Note that the father is heterozygous at locus A and homozygous at locus B. The alleles in the child will be a combination of one allele from each parent.

sizes of the offspring and the mother are analyzed. The remaining fragments (the ones that do not match the mother) have to come from the father. Alleged fathers are identified based on the ability to provide the remaining alleles **(inclusion).** Aside from possible mutations, a difference in just one allele may exclude paternity.

A simplified RFLP paternity test is shown in Figure 10.5. Of the two alleged fathers shown, only one could supply the fragments not supplied by the mother. In this example, only two loci are shown. A parentage test requires analysis of at least eight loci. The more loci tested, the higher the probability of positive identification of the father.

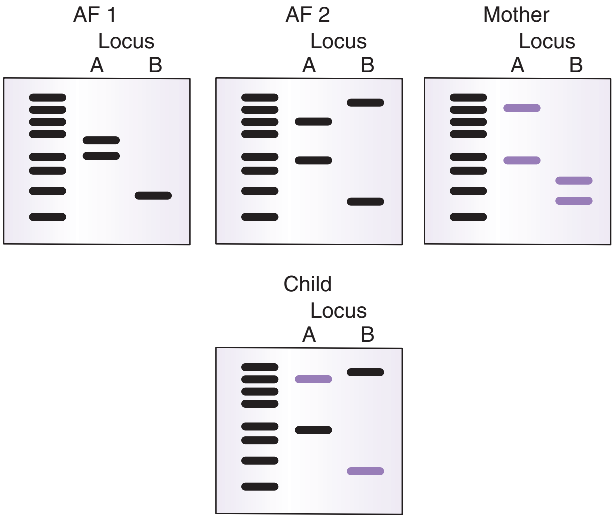


FIGURE 10.5 Two alleged fathers (AFs) are being tested for paternity of the child whose partial RFLP profile is shown in the bottom gel. The mother's alleles are shown in green. One AF (AF1) is excluded from paternity because he cannot supply the child's paternal allele at locus B.

Human Identification Using RFLPs

The first genetic tool used for human identification was the ABO blood group antigens. Although this type of analysis could be performed in a few minutes, the discrimination power was low. With only four possible groups, this method was only good for **exclusion** (elimination) of a person and was informative only in 15% to 20% of cases. Analysis of the polymorphic HLA loci added a higher level of discrimination, with exclusion in 90% of cases. Testing both ABO and HLA did not provide positive identification, however.

The initial use of DNA as an identification tool relied on RFLP detectable by Southern blot. As shown in Figure 10.1, RFLP can arise from a number of genetic events, including point mutations in the restriction site, mutations that create a new restriction site, and insertion or deletion of repeated sequences (tandem repeats). The insertion or deletion of nucleotides occurs frequently in repeated sequences in DNA. Tandem repeats of sequences of all sizes are present in genomic DNA (Fig. 10.6). Repeat units can be large enough so that loss or gain of one repeat is resolved by gel electrophoresis of a restriction enzyme digest. The frequent cutters, *Hae*III (recognition site GGCC) or *Hinf*I (recognition site GANTC), generate fragments that are small enough to resolve those that contain different numbers of repeats and thereby give an informative pattern by Southern blot.

DNA Fingerprinting With RFLP

The first human DNA profiling system was introduced by the United Kingdom Forensic Science Service in 1985 using Sir Alec Jeffreys's Southern blot **multiple-locus probe (MLP)**-RFLP system.[5] This method utilized three to five probes to analyze three to five loci on the same

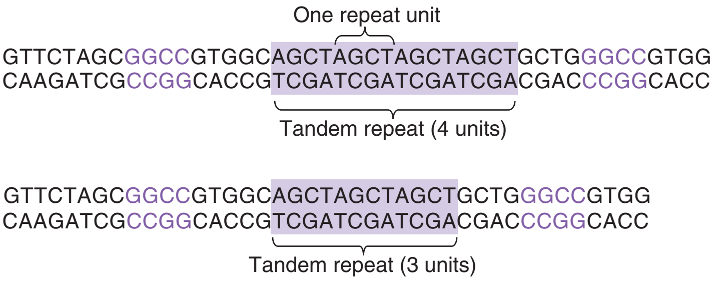

One repeat unit
GTTCTAGCGGCCGTGGCAGCTAGCTAGCTAGCTGCTGGGCCGTGG
CAAGATCGCCGGCACCGTCGATCGATCGATCGACGACCCGGCACC
Tandem repeat (4 units)
GTTCTAGCGGCCGTGGCAGCTAGCTAGCTGCTGGGCCGTGG
CAAGATCGCCGGCACCGTCGATCGATCGACGACCCGGCACC
Tandem repeat (3 units)

FIGURE 10.6 A tandem repeat is a direct repeat of 1 to more than 100 nucleotides in length. The one shown has a 4-bp repeat unit (AGCT). A gain or loss of repeat units forms a different allele. Different alleles are detected as variations in fragment size on digestion with a restriction enzyme, such as *Hae*III (GGCC recognition sites).

blot. Results of probing multiple loci at once produced patterns that were highly variable between individuals but that required some expertise to optimize and interpret. In 1990, **single-locus probe (SLP)** systems were established in Europe and North America.[6,7] Analysis of one locus at a time yielded simpler patterns, which were much easier to interpret, especially in cases where specimens might contain a mixture of DNA from more than one individual (Fig. 10.7).

The RFLP Southern blot technique required 100 ng to 1 µg of relatively high-quality DNA, 1 to 20 kbp in size. Furthermore, large, fragile 0.7% gels were required to achieve adequate band resolution, and the ^{32}P-based probe system could take 5 to 7 days to yield clear results. After visually inspecting the band patterns, profiles were subjected to computer analysis to accurately size the restriction fragments and apply the results to an established matching criterion. RFLP is an example of a **continuous allele system** in which the sizes of the fragments define alleles. Therefore, precise band sizing was critical to the accuracy of the results. A match implied inclusion, which was refined by determination of the genotype frequency of each allele in the general or local population. This process established the likelihood of the same genotype occurring by chance. The probability of two people having the same set of RFLP, or **profile,** becomes lower and lower as more loci are analyzed.

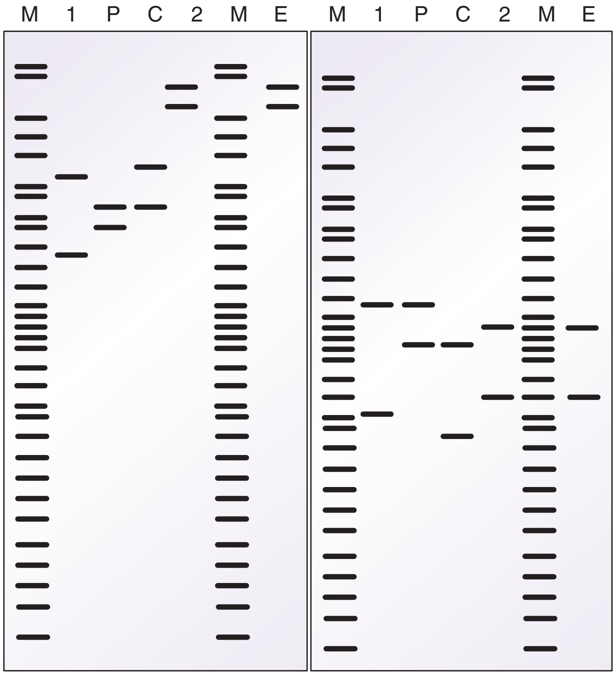


FIGURE 10.7 Example of RFLP crime evidence using two single-locus probes. M denotes molecular-weight markers, 1 and 2 are suspects, C is the child victim, and P is the parent of the child victim. E is evidence from the crime scene. For both loci probed, suspect 2 "matches" the evidence found at the crime scene. Positive identification of suspect 2 requires further determination of the frequencies of these specific alleles in the population and the probability of matching them by chance.

Historical Highlights

Professor Sir Alec John Jeffreys, a British geneticist, first developed techniques for genetic profiling, or DNA fingerprinting, using RFLP to identify humans. The technique has been used in forensics and law enforcement to resolve paternity and immigration disputes. The method can also be applied to nonhuman species, for example, in wildlife population genetics. The first application of this DNA technique was in a regional screen of human DNA to identify the rapist and killer of two girls in Leicestershire, England, in 1983 and 1986. Colin Pitchfork was identified and convicted of murder after samples taken from him matched semen samples taken from the two victims.

STR TYPING BY PCR

The first commercial and validated typing test based on polymerase chain reaction (PCR) specifically for forensic use was the HLA DQ alpha system, now called DQA1, developed in 1986.[8] This system could distinguish 28 DQA1 types. With the addition of another commercial system, the Polymarker (PM) system, the analyst could type five additional genetic markers. The PM system is a set of primers complementary to sequences flanking STRs, or **microsatellites.** STRs are similar to VNTRs **(minisatellites)** but have repeat units of 1 to 7 bp. (The upper limit of repeat unit size for STR varies from 7 to 10 bp, depending on different texts and reports.) Because of the increased power of discrimination and ease of use of STR, the HLA DQA forensic DNA amplification and typing kit was discontinued in 2002.

The tandem repeat shown in Figure 10.6 is an STR with a 4-bp repeat unit, AGCT. Occasionally, STRs contain repeat units with altered sequences, or **microvariants,** repeat units missing one or more bases of the repeat. These differences have arisen through mutation or recombination events.

In contrast to VNTRs, the smaller STRs are efficiently amplified by PCR, easing specimen demands significantly. Long, intact DNA fragments are not required to detect the STR products; therefore, degraded or otherwise less-than-optimal specimens are potentially informative. The amount of specimen required for STR analysis by PCR is reduced from 1 µg to 10 ng, a key factor for forensic analysis. Furthermore, PCR procedures shorten the analysis time from several weeks to 24 to 48 hours. Careful design of primers and amplifications facilitated multiplexing and automation of the process.[9]

Advanced Concepts

Although STRs with 4- and 5-bp repeat units are highly informative and efficiently amplified, they are subject to naturally occurring genetic events. Loss or gain of repeats or parts of repeat units, as well as mutations within repeat units, are very rare occurrences. Because at least 8 to more than 20 loci are included in STR applications, these minor events do not significantly affect the informative power. (Allele population frequency has the most limiting effect on inclusion.) In abnormal cells with genetic instability, such as cancer cells, gain or loss of repeats can occur more frequently, enough to affect the identification of genotypes.[10]

STR alleles are identified by PCR product size. Primers are designed to produce amplicons of 100 to 400 bp in which the STRs are embedded (Fig. 10.8). The sizes of the PCR products are influenced by the number of embedded repeats. If one of each primer pair is labeled with a fluorescent marker, the PCR product can be analyzed in fluorescent detection systems. Silver-stained gels may also be used; however, capillary electrophoresis with fluorescent dyes is the preferable method, especially for high-throughput requirements.

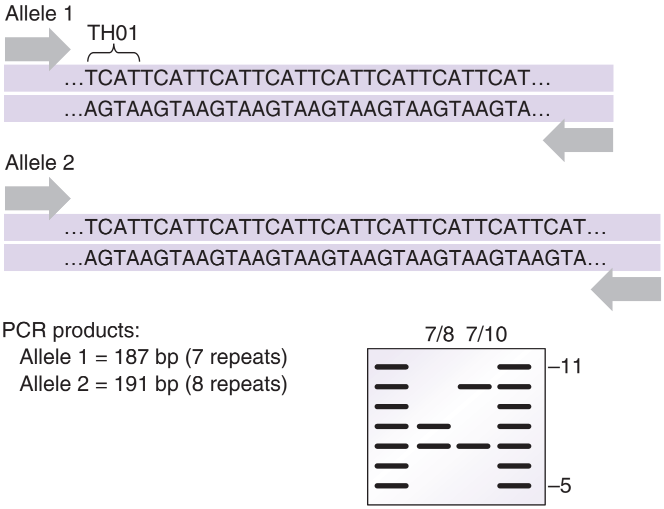

Allele 1
TH01
...TCATTCATTCATTCATTCATTCATTCATTCAT...
...AGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTA...
Allele 2
...TCATTCATTCATTCATTCATTCATTCATTCATTCAT...
...AGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTA...
PCR products:
Allele 1 = 187 bp (7 repeats)
Allele 2 = 191 bp (8 repeats)
7/8 7/10
–11
–5

FIGURE 10.8 STR TH01 (repeat unit TCAT) linked to the human tyrosine hydroxylase gene on chromosome 11p15.5. Primers are designed to amplify short regions containing the tandem repeats. Allelic ladders consisting of all alleles in the human population (flanking lanes in the gel shown at bottom right) are used to determine the number of repeats in the locus by the size of the amplicon. The two alleles shown contain seven and eight repeats. If these alleles were found in a single individual, that person would be heterozygous for TH01 with a genotype of 7/8. Compare the 7/8 genotype pattern with the 7/10 genotype gel pattern.

Historical Highlights

At least three to seven RFLP probes were originally required to determine genetic identity. Available probes included G3, MS1, MS8, MS31, and MS43, which were subclones of Jeffreys's multilocus probes 33.6 and 33.15 and pYNH24m, MS205, and MS621. SLPs MS1, MS31, MS43, G3, and YNH24 were used in the O. J. Simpson trial in 1996.

A further development of STR analysis was the design of **mini-STR.** These STRs are amplified with PCR primers located closer to the tandem repeat than in the standard STR. Compared with standard STR products, the small amplicons are more efficiently produced from such challenging starting material as fixed tissue[11] and degraded specimens.[12] Another specialized system, **Y-STR,** was developed for surname testing and forensic identification of male offenders or victims. This primer set only amplifies STR located on the Y chromosome. There is only one allele at each locus, and because the Y chromosome is inherited as a single haplotype, paternally related men share all Y loci.[13]

STR Analysis

To identify STR alleles, test DNA is mixed with the primer pairs, buffer, and polymerase to amplify the test loci. Primer pairs may be laboratory designed or purchased commercially. A control DNA standard is also amplified, as well as a sensitivity control, if the relative allele percentage in a mixture will be calculated. Following amplification, each sample PCR product is combined with **allelic ladders** (sets of fragments representing all possible alleles of a repeat locus) and internal size standards (molecular-weight markers) in formamide for electrophoresis. After electrophoresis, detection and analysis software will size and identify the alleles based on co-size migration with specific alleles in the allelic ladders. In contrast to RFLPs and VNTRs, STRs are **discrete allele systems** in which a finite number of alleles is defined by the number of repeat units in the tandem repeat (see Fig. 10.8). Several available commercial systems consist of labeled primers for 1 to more than 16 loci. The allelic ladders in these reagent kits allow accurate identification of the sample alleles (Fig. 10.9).[14]

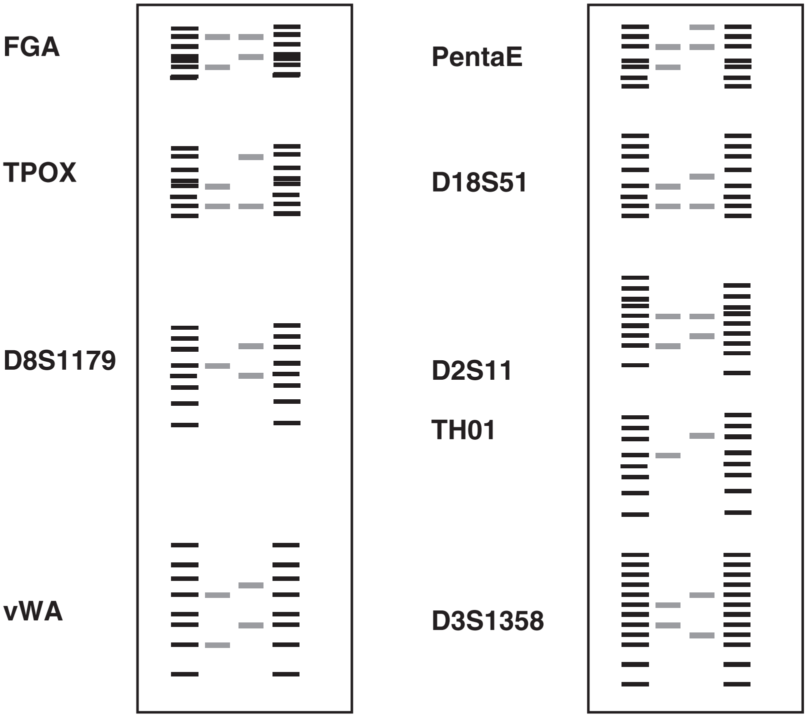


FIGURE 10.9 Multiple STRs can be resolved on a single gel. Here, four and five different loci are shown on the left and right gels, respectively. The allelic ladders show that the ranges of potential amplicon sizes do not overlap, allowing resolution of multiple loci in the same lane. Two individual genotypes are shown on the second and third lanes of the two gels.

Advanced Concepts

Theoretically, the minimal sample requirement for PCR analysis is a single cell. A single cell has approximately 6 pg of DNA. This number is derived from the molecular weight of A/T and G/C base pairs (617 and 618 g/mol, respectively). There are about 3 billion base pairs in one copy of the human genome; therefore, for one genome copy:

$$3 \times 10^{9}\ \text{bp} \times 618\ \text{g/mol/bp} = 1.85 \times 10^{12}\ \text{g/mol}$$

$$1.85 \times 10^{12}\ \text{g/mol} \times 1\ \text{mol}/6.023 \times 10^{23}\ \text{molecules} = 3.07 \times 10^{-12}\ \text{g} = 3\ \text{pg}$$

(A diploid cell has two genome copies, or 6 pg of DNA.) One ng (1,000 pg) of DNA should, therefore, contain 333 copies (1,000 pg/3 pg/genome copy) of each locus.

Advances in fluorescence technology have increased the ease and sensitivity of STR allele identification (Fig. 10.10). Although capillary electrophoresis is faster and more automated than gel electrophoresis, a single run through a capillary of single dye-labeled products can resolve only loci whose allele ranges do not overlap. The number of loci that can be resolved on a single run was increased by the use of multicolor dye labels. Primer sets labeled with dyes that can be distinguished by their emission wavelength generate products that are resolved according to fluorescent color as well as size (Fig. 10.11). Test DNA amplicons, allelic ladders, and size standards for multiple loci are thus run simultaneously through each capillary. Genotyping software provides automated resolution of fluorescent dye colors and genotyping by comparison with the size standards and the allelic ladder.

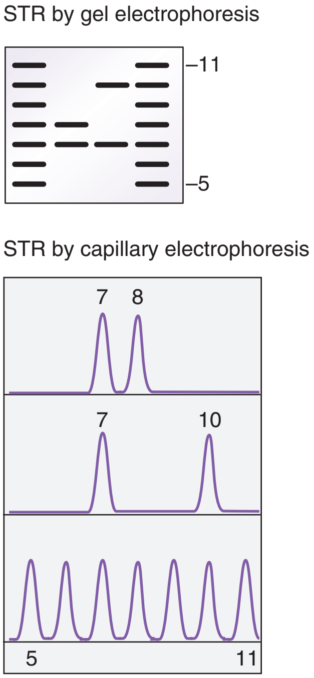


FIGURE 10.10 STR analysis by capillary gel electrophoresis. Instead of bands on a gel *(top)*, peaks of fluorescence on an electropherogram reveal the PCR product sizes *(bottom)*. Alleles (7, 8, or 10) are determined by comparison with allelic ladders representing all possible alleles (from 5 to 11 repeats) for this locus, run through the capillary simultaneously with the sample amplicons.

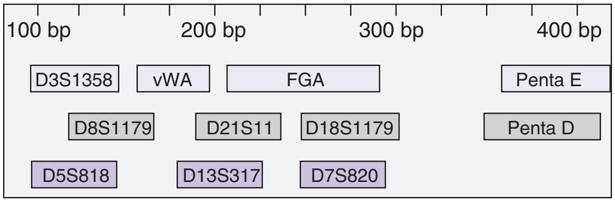


FIGURE 10.11 An illustration of the ranges of allele peak locations for selected STRs. By labeling primers with different fluorescent dye colors, STRs with overlapping size ranges can be resolved by color. The molecular-weight markers *(bottom)* are labeled with a fluorescent dye distinguishable from those used for the primer labeling.

Advanced Concepts

Commercial primer sets are designed with "stuffer" sequences to modify the size of the PCR products so that the range of alleles for four to five loci can be resolved by electrophoresis. Therefore, the product size for a given allele will not always be the same with primers from different commercial sources.

As in RFLP testing, an STR "match" is made by comparing profiles (alleles at all loci tested) followed by probability calculations. The HLA DQ in conjunction with the PM system generated highly discriminatory allele frequencies. For example, the chance of the same set of alleles or profile occurring in two unrelated individuals at random is 1 in 10^6 to 7×10^8 Caucasians or 1 in 3×10^6 to 3×10^8 African Americans.

STR Nomenclature

The International Society for Forensic Genetics recommended nomenclature for STR loci in 1997.[15] STRs within genes are designated according to the gene name; for example, the STR TH01 is located in intron 1 of the human tyrosine hydroxylase gene on chromosome 11, and TPOX is located in intron 10 of the human thyroid peroxidase gene on chromosome 2. These STRs do not have any phenotypic effect with respect to these genes. Non-gene-associated STRs are designated by the D#S# system. In this system, the "D" stands for DNA; the following number designates the chromosome where the STR is located (1-22, X or Y). "S" refers to a unique segment, followed by a number registered in the International Genome Database (GDB). See Table 10.2 for some examples.

STRs are present all over the genome. Some of the STR loci commonly used for laboratory investigation are shown in Table 10.2. A comprehensive collection of STR information is available at http://www.cstl.nist.gov/strbase/.

Gender Identification

The **amelogenin locus** is a very useful marker often analyzed along with STR. The amelogenin gene, which is not an STR, is located on the X and Y chromosomes. The function of its encoded protein is required for embryonic development and tooth maturation. A polymorphism is located in the second intron of the amelogenin gene. The Y allele of the gene is 6 bp larger in this region than in the X allele. Amplification and electrophoretic resolution reveal two bands or peaks for males (XY) and one band or peak for females (XX; Fig. 10.12). Some commercially available sets will contain primers to amplify the amelogenin polymorphism in addition to the STR primer sets. Additional loci are now available for gender testing, in cases where amelogenin may be compromised or poorly amplified.[16]

Analysis of Test Results

Analysis of polymorphisms at multiple loci results in very high levels of discrimination (Table 10.3). Discovery of the same set of alleles from different sources or shared alleles between allegedly related individuals is strong evidence of identity, paternity, or relatedness. Results from such studies, however, must be expressed in terms of the background probability of chance matches.

Historical Highlights

The GDB is overseen by the Human Genome Nomenclature Committee, a part of the Human Genome Organization (HUGO) located at University College, London. HUGO was established in 1989 as an international association of scientists involved in human genetics. The goal of HUGO is to promote and sustain international collaboration in the field of human genetics. The GDB was originally used to organize mapping data during the earliest days of the Human Genome Project. With the release of the human genome sequences and the development of PCR, the number of laboratories doing genetic testing grew significantly. The GDB is still widely used as a source of information about PCR primers, PCR products, polymorphisms, and genetic testing. Information from the GDB is available at http://www.ncbi.nlm.nih.gov/sites/genome.

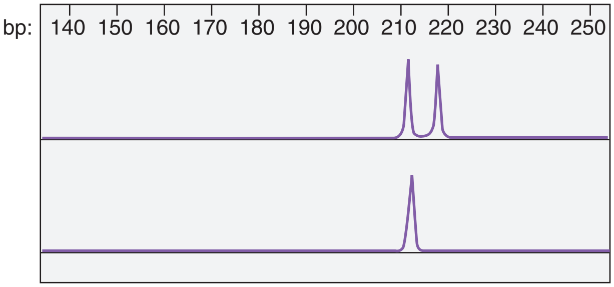


FIGURE 10.12 Amplification of amelogenin will produce a male-specific 218-bp product (Y allele) in addition to the 212-bp product found on the X chromosome (X allele). Males are heterozygous for the amelogenin locus (XY, *top*), and females are homozygous for this locus (XX, *bottom*).

TABLE 10.2 STR Locus Information*[71]

STR	Locus	Chromosome Sequence	Repeat	Alleles[†]
CD4	Locus between CD4 and triosephosphate isomerase	12p	AAAAG[‡]	4, 6, 7, 8, 8′, 9, 10, 11, 12, 13, 14, 15
CSF1PO	c-fms protooncogene for CSF-1 receptor	5q	TAGA	6, 7, 8, 9, 10, 11, 12, 13, 14, 15
D3S1358		3p	TCTA[§]	8, 9, 10, 11, 12, 13, 14, 15, 15′, 15.2, 16, 16′, 16.2, 17, 17′, 17.1, 18, 18.3, 19, 20
D5S818		5q	AGAT	7, 8, 9, 10, 11, 12, 13, 14
D7S829		7q	GATA	7, 8, 9, 10, 11, 12, 13, 14, 15
D8S1179	Sequence tagged site	8q	TCTA	7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19
D13S317		13q	TATC	7, 8, 9, 10, 11, 12, 13, 14, 15
D16S539		16q	GATA	5, 8, 9, 10, 11, 12, 13, 14, 15
D18S51	Sequenced tagged site	18q	GAAA	7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27
D21S11[‖]		21q	TCTG	24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38
F13A01	Coagulation factor IX	6p	GAAA	3.2, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16
F13B	Factor XIII b	1q	TTTA	6, 7, 8, 9, 10, 11, 12
FESFPS	c-fes/fps protooncogene	15q	ATTT	7, 8, 9, 10, 11, 12, 13, 14
HPRTB	Hypoxanthine phosphoribosyl-transferase	Xq	TCTA	6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17
LPL	Lipoprotein lipase	8p	TTTA	7, 8, 9, 10, 11, 12, 13, 14
TH01	Tyrosine hydroxylase	11p	TCAT	5, 6, 7, 8, 9, 9.3, 10, 11
TPOX	Thyroid peroxidase	2p	TGAA	6, 7, 8, 9, 10, 11, 12, 13
vWA	Von Willebrand factor	12p	TCTA	11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21
Penta D		21q	AAAGA	2.2, 3.2, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17
Penta E		15q	AAAGA	5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 20.3, 21, 22, 23, 24

*http://www.cstl.nist.gov/

[†] Some alleles have units with one, two, or three missing bases.

[‡] In an alternate 8-repeat allele, one repeat sequence is AAAGG.

[§] In alternate 15-, 16-, or 17-repeat alleles, one repeat sequence is TCTG.

[‖] D21S11 has multiple alternate alleles.

TABLE 10.3 Matching Probability of STR Genotypes in Different Subpopulations

	African American	White American	Hispanic American
8 loci	1/274,000,000	1/114,000,000	1/145,000,000
9 loci	$1/5.18 \times 10^{9}$	$1/1.03 \times 10^{9}$	$1/1.84 \times 10^{9}$
10 loci	$1/6.76 \times 10^{10}$	$1/9.61 \times 10^{10}$	
12 loci	$1/4.61 \times 10^{12}$		
14 loci	$1/6.11 \times 10^{17}$	$1/9.96 \times 10^{17}$	$1/1.31 \times 10^{17}$
16 loci	$1/7.64 \times 10^{17}$	$1/9.96 \times 10^{17}$	$1/1.31 \times 10^{17}$

Advanced Concepts

In 1997 the Federal Bureau of Investigation adopted 13 "core" loci as the **Combined DNA Indexing System (CODIS).** The loci are TPOX on chromosome 2, D3S1358 on chromosome 3, FGA on chromosome 4, D5S818 and CSF1PO on chromosome 5, D7S820 on chromosome 7, D8S1179 on chromosome 8, TH01 on chromosome 11, vWA on chromosome 12, D13S317 on chromosome 13, D16S539 on chromosome 16, D18S51 on chromosome 18, D21S11 on chromosome 2, and the amelogenin locus on the X and Y chromosomes. The National Institute of Standards and Technology supplies Standard Reference Material that certifies values for 22 STR loci, including CODIS and markers used by European forensic laboratories.

Genotyping

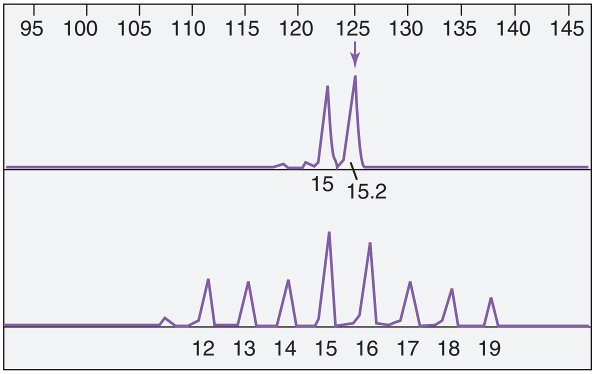


FIGURE 10.13 A microvariant allele (15.2, *top*) migrates between the full-length alleles (15 and 16 on the allelic ladder, *bottom*).

DNA testing results in peak or band patterns that must be converted to genotype (allele identification). As described previously, an STR locus genotype is defined by the number of repeats in its alleles. For instance, if the **locus genotype** in Figure 10.8 represented homologous chromosomes from an individual, the locus would be heterozygous, with 7 repeats on one chromosome and 8 repeats on the other. This locus would thus be designated 7/8 or 7,8. A homozygous locus (where both homologous chromosomes carry the same allele) is designated by the 7/7 or 7,7. Some reports use a single number, such as 6 or 7, to designate a homozygous locus. Microvariant alleles containing partial repeat units are indicated by the number of complete repeats followed by a decimal point and then the number of bases in the partial repeat. For example, the 9.3 allele of the TH01 locus has 10 repeats, 9 full 4-bp repeat units and 1 repeat unit with 3 bp. Microvariants are detected as bands or peaks very close to the full-length allele (Fig. 10.13).

The genotype, or profile, of a specimen is the collection of alleles in all the loci tested. To determine the extent of certainty that one profile matches another, the occurrence of the detected genotype in the general or a defined population must be assessed.

A matching genotype is not necessarily an absolute determination of the identity of an individual. **Genetic concordance** is a term used to express the situation where all locus genotypes (alleles) from two sources are the same. Concordance is interpreted as inclusion of a single individual as the donor of both genotypes. Two samples are considered different if at least one locus genotype differs (exclusion). An exception is paternity testing, in which mutational events may generate a new allele in the offspring at one locus, and this difference may not rule out paternity.

Technical artifacts such as air bubbles, crystals, and dye blobs, as well as sample contaminants, temperature variations, and voltage spikes, can interfere with consistent band migration during electrophoresis. In addition, amplification artifacts occur during PCR. Some polymerases add an additional non-template adenine residue to the 3′ end of the PCR product. If this 3′ nucleotide addition does not include all the amplified products, a mixed set of amplicons will result in extra bands or peaks located very close together.

Stutter is another anomaly of PCR amplification, in which the polymerase may miss a repeat during the replication process, resulting in two or more different species in the amplified product. These also appear as extra bands or peaks. Generally, the larger the repeat unit length, the less stutter is observed. These or other aberrant band patterns confuse the analysis software and can result in the miscalling of alleles.

Matching requires clear and unambiguous laboratory results. As alleles are identified by gel resolution, good intragel precision (comparing bands or peaks on the same gel or capillary) and intergel precision (comparing bands or peaks of separate gels or capillaries) are important. In general, intergel precision is less stringent than intragel precision. This is not unexpected because the same samples may run with slightly different migration speeds on different gels. Because some microvariant alleles differ by only a single base pair (see Fig. 10.13), the resolution must be less than ±0.5 bp. The TH01 9.3 allele described earlier is an example. This allele must be distinguished from the 10 allele, which is a single base pair larger than the 9.3 allele.

To establish the identity of peaks from capillary electrophoresis (or peaks from densitometry tracings of gel data), the peak is assigned a position relative to some landmark within the gel lane or capillary, such as the loading well or the start of migration. Upon replicate resolutions of a band or peak, electrophoretic variations from capillary to capillary, lane to lane, or gel to gel may occur. Normalization of migration is achieved by the relation of the migration of the test peaks to the simultaneous migration of size standards. Size standards can be internal (in the same gel lane or capillary) or external (in a separate gel lane). Even with normalization, however, tiny variations in position, height, and area of peaks or gel bands may persist. If the same fragments are run repeatedly, a distribution of observed sizes can be established. An acceptable range of sizes in this distribution is a **bin.** A bin can be thought of as an uncertainty window surrounding the mean position (size) of multiple runs of each peak or band. All bands or peaks, therefore, that fall within this window are considered identical. Collection of all peaks or bands within a characteristic distribution of positions and areas is called **binning.** Bins for each allele can be established manually in the laboratory. Alternatively, commercially available software has been designed to automatically bin and identify alleles.[17]

All peaks within a bin are interpreted as representative of the same allele of a locus. Each band or peak in a genotype is binned and identified according to its migration characteristics. The group of bands or peaks makes up the characteristic pattern or profile of the specimen.

Advanced Concepts

Binning may be performed in different ways using replicate peak heights and positions. To calculate the probability that two peaks are representative of the same allele, the proportion of alleles that fall within the uncertainty window (bin) must be determined. This proportion is represented exactly in fixed bins and approximated in floating bins. The fixed-bin approach is an approximation of the more conservative floating-bin approach.[18] An alternative assessment of allele certainty is the use of **locus-specific brackets.** In this approach, artificial “alleles” are designed to run at the high and low limit of the expected allele size. Identical alleles are expected to fall within this defined bracket.[17]

Matching of Profiles

The number of loci tested must be taken into consideration in genotyping analysis. The more loci analyzed, the higher the probability that the locus genotype positively identifies an individual (**match probability;** see Table 10.3). Degraded, compromised, or mixed samples will affect the match probability because all loci may not yield clear, informative results. Criteria for interpretation of results and determination of a true allele are established by each laboratory. These criteria are based on validation studies and results reported from other laboratories. Periodic external proficiency testing is performed to confirm the accuracy of test performance.

Results from the analysis of polymorphisms are used to determine the probability of identity or inheritance of genetic markers or to match a particular marker or marker pattern. To establish the identity of an individual by an allele of a locus, the chance that the same allele could arise in the population randomly is taken into account.

Advanced Concepts

The certainty of a matching pattern increases with decreased frequency of alleles in the general population. Under defined conditions, the relative frequency of two alleles in a population remains constant. This is **Hardy–Weinberg equilibrium,** or the Hardy–Weinberg law.[19] The population frequency of two alleles, p and q, can be expressed mathematically as

$$p^2 + 2pq + q^2 = 1.0$$

This equilibrium assumes a large population with random mating and no immigration, emigration, mutation, or natural selection. Under these circumstances, if enough individuals are assessed, a close approximation of the true allele frequency in the population can be determined.

The frequency of a set of alleles or a genotype in a population is the product of the frequency of each allele separately (the **product rule**). The product rule can be applied because of linkage equilibrium. Linkage equilibrium assumes that the observed frequencies of haplotypes in a population are the same as haplotype frequencies predicted by multiplying together the frequency of individual genetic markers in each haplotype and that loci are not genetically linked (located close to one another) in the genome. The overall frequency (OF) of a locus genotype consisting of *n* loci can be calculated as

$$OF = F_1 \times F_2 \times F_3 \times \ldots F_n$$

where $F_{1\,...\,n}$ represents the frequency of each individual allele in the population.

Individual allele frequencies are determined by data collected from testing many individuals in general and defined populations. For example, at locus Penta D on chromosome 21, the 5 allele has been previously determined to occur in 1 of 10 people in a theoretical population. At locus D7S829 on chromosome 7, the 8 allele has been previously observed in 1 of 50 people in the same population. The overall frequency of the profile containing the loci Penta D 5 allele and D7S829 8 allele would be 1/10 × 1/50 = 1/500. That is, a genotype or profile containing D7S829 8 and Penta D 5 alleles would be expected to occur in 1 out of every 500 randomly chosen members of that population. As should be apparent, the more loci tested, the greater the certainty that the profile is unique to a single individual in that population; that is, the overall frequency of the profile is very low. The overall frequencies in Table 10.3 illustrate this point.

Allele frequencies differ between **subpopulations** or ethnic groups. Different allele frequencies in subpopulations are determined through the study of each ethnic group.[20] The data in Table 10.3 illustrate differences in the polymorphic nature of alleles in different subpopulations.

When profile identification requires comparing the genotype of an unknown specimen with a known reference sample—for example, the genotype of evidence from a crime and the genotype of an individual from a database—the determination that the two genotypes match (are from the same person) is expressed in terms of a **likelihood ratio.** The likelihood ratio is the comparison of the probability that the two genotypes came from the same person with the probability that the two genotypes came from different persons, taking into account allele frequencies and **linkage equilibrium** in the population. A high likelihood ratio is an indication that the probability is more likely that the two genotypes came from the same person, whereas a likelihood ratio of less than 1 indicates that this probability is less likely. If a likelihood ratio is [1/(1/1,000)] = 1,000, the tested genotypes are 1,000 times more likely to have come from the same person than from two randomly chosen members of the population, where the profile occurs in 1/1,000 people. In a random sampling of 100,000 members of a population, 100 people (100,000/1,000) with the same genotype might be found.

A simplified illustration can be made from the previous Penta D and D7S829 example. Suppose the Penta

D 5 (1/10 allele frequency) and D7S829 8 (1/50 allele frequency) profiles were discovered in a specimen from an independent source. The likelihood that the Penta D 5/D7S829 8 profile came from the tested individual is 1, having been directly determined. The likelihood that the same profile could come from someone else in the population is 1/500. The likelihood ratio is 1/(1/500), or 500. The specimen material is 500 times more likely to have come from the tested individual than from some other person in the population.

Historical Highlights

Sir Alec Jeffreys' DNA profiling was the basis for the National DNA Database (NDNAD) launched in Britain in 1995. Under British law, the DNA profile of anyone convicted of a serious crime is stored on a database. This database includes 6% of the UK population (compared with 0.5% of the population in the U.S. database). Over 5.9 million DNA profiles are held in the database, accounting for a majority of the known active-offender population. Nearly 60% of crime-scene profiles submitted to the NDNAD were matched to a subject profile between 2008 and 2009.

The National DNA Index System (NDIS) is the federal level of the CODIS used in the United States. There are three levels of CODIS: the Local DNA Index System (LDIS), State DNA Index System (SDIS), and NDIS. At the local level, CODIS software maintained by the Federal Bureau of Investigation (FBI) is used in sizing alleles as the assay is performed. This information may be compiled locally and/or submitted to the SDIS. The state data may be sent to the NDIS. The SDIS and NDIS must adhere to the quality assurance standards recommended by the FBI. The original entries to these databases were RFLP profiles; further entries have been the STR profiles. As of October 2018, the NDIS contained over 13 million offender profiles.

When comparing genotypes with those in a database looking for a match, it is important to consider whether the database is representative of a population or subpopulation because allele frequencies will be different in differently defined groups. It is also important to consider whether the population is homogeneous (a random mixture) with respect to the alleles tested. Familial searches and forensic applications involving mass disasters or other complex mixtures of DNA involve analysis of partial or uncertain DNA genotypes. More advanced approaches, such as "wild card" designations for missing alleles, are required for defining the likelihood in these cases.[21]

Allelic Frequencies in Paternity Testing

A paternity test is designed to choose between two hypotheses: The test subject is not the father of the tested child (H_0), or the test subject is the father of the tested child (H_1). Paternity is first assessed by observation of shared alleles between the alleged father and the child (Fig. 10.14). The identity of shared alleles is a process of matching, as described previously for identity testing.

Historical Highlights

Peter Gill developed a forensic DNA identification method for minimal samples called low-copy-number analysis.[22] In contrast to standard DNA analysis that requires approximately 200 pg DNA, low-copy-number analysis is reportedly performed on less than 100 pg DNA (about 16 diploid cells). The method involves increasing the number of PCR cycles for amplification and fastidiously cleaned work areas. Although used in over 21,000 serious criminal cases since the 1990s, the validity of this technique has been questioned in appeal cases.[23] Assaying limited amounts of starting material may result in peak dropouts (failed amplification of an allele) or drop-ins (mis-priming of an allele). Furthermore, the heightened detection limit required for low amounts of target DNA raises the risk of contamination.

Because each allele of a genotype is inherited from one parent, a child will share one allele of every locus with the paternal parent. A **paternity index,** or likelihood of paternity, is calculated for each locus in which the alleged father and the child share an allele. The paternity

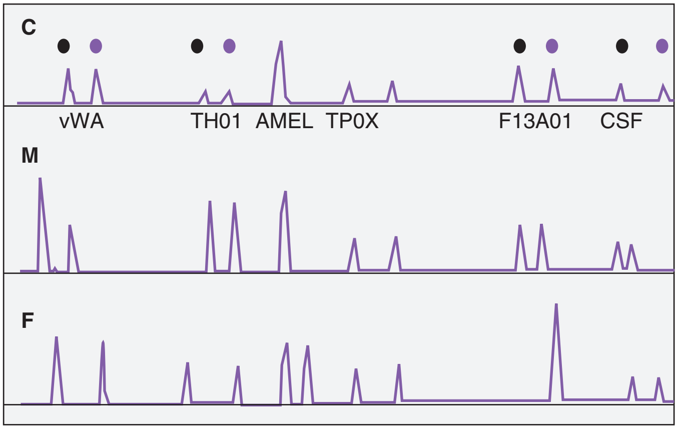


FIGURE 10.14 Electropherogram showing results from five STR loci and the amelogenin locus for a child (C), mother (M), and father (F). Note how the child has inherited one of each allele from the mother (black dots) and one from the father.

TABLE 10.4 Example Data From a Paternity Test Showing Inclusion

Alleged	Child	Father Allele	Shared Allele	Paternity Index
D16S539	8, 9	9, 10	9	5.719
D5S818	10, 12	7, 12	12	8.932
FESFPS	9, 13	13, 14	13	15.41
F13A01	4, 5	5, 7	5	10.22

index is an expression of how many times more likely the child's allele is inherited from the alleged father than by another man in the general population. An allele that occurs frequently in the population has a low paternity index; a rare allele has a high paternity index. Table 10.4 shows the paternity index for each of four loci. The FESFPS 13 allele is rarer than the D 16S539 9 allele. In this example, the child is 5.719 times more likely to have inherited the 9 allele of locus D16S539 from the alleged father than from another random man in the population. Similarly, the child is 15.41 times more likely to have inherited the 13 allele of FESFPS from the alleged father than by random occurrence. When each tested locus is on a different chromosome (not linked), the inheritance or occurrence of each allele can be considered an independent event. The paternity index for each locus, therefore, can be multiplied together to calculate the **combined paternity index (CPI),** which summarizes and evaluates the genotype information. The CPI for the data shown in Table 10.4 are

$$\text{CPI} = 5.719 \times 8.932 \times 15.41 \times 10.22 = 8{,}044.931$$

These data indicate that the child is 8,045 times more likely to have inherited the four observed alleles from the alleged father than from another man in the population.

If a paternal allele does not match between the alleged father and the child, H_1 for that allele is 0. One might assume, therefore, that the nonmatching allele paternity index of 0 would make the CPI 0. This is not the case. Nonmatching alleles between the alleged father and the child found at one locus (exclusion) is traditionally not regarded as a demonstration of non-paternity because of the possibility of mutation. Although mutations were quite rare in the traditional RFLP systems, analysis of 12 or more STR loci may occasionally reveal one or two mutations resulting in nonmatching alleles even if the man is the father. To account for mutations, the paternity index for nonmatching alleles is calculated as

$$\text{paternity index for a mutant allele} = \mu$$

where μ is the observed mutation rate (mutations/meiosis) of the locus. The American Association of Blood Banks (AABB) has collected data on mutation rates in STR loci (Table 10.5). Using these data, in the case of a nonmatching allele, H_1 is not 0 but μ. A high mutation rate (close to 1) would not lower the CPI, whereas a very low mutation rate (closer to 0) would do so.

In a paternity report, the combined paternity index is accompanied by the **probability of paternity,** a number calculated from the combined paternity index (genetic evidence) and **prior odds** (nongenetic evidence). For the

TABLE 10.5 Observed Mutation Rates in Paternity Tests Using STR Loci

STR Locus	Mutation Rate (%)
D1S1338	0.09
D3S1358	0.13
D5S818	0.12
D7S820	0.10
D8S1179	0.13
D13S317	0.15
D16S539	0.11
D18S51	0.25
D19S433	0.11
D21S11	0.21
CSF1PO	0.16
FGA	0.30
TH01	0.01
TPOX	0.01
VWA	0.16
F13A01	0.05
FESFPS	0.05
F13B	0.03
LPL	0.05
Penta D	0.13
Penta E	0.16

prior odds, the laboratory as a neutral party assumes a 50/50 chance that the test subject is the father. Therefore, the probability of paternity is

$$\frac{\text{CPI} \times \text{prior odds}}{(\text{CPI} \times \text{prior odds}) + (1 - \text{prior odds})}$$

$$= \frac{\text{CPI} \times 0.50}{(\text{CPI} \times 0.50) + (1 - 0.50)}$$

TABLE 10.6 Odds of Paternity Using Different Prior Odds Assumptions

Prior Odds					
CPI	10%	25%	50%	75%	90%
5	0.36	0.63	0.83	0.94	0.98
9	0.50	0.75	0.90	0.96	0.98
19	0.68	0.86	0.95	0.98	0.994
99	0.92	0.97	0.99	0.997	0.999
999	0.99	0.997	0.999	0.9997	0.9999

In the example illustrated previously, the CPI is 8,044.931. The probability of paternity is

$$\frac{8{,}044.931 \times 0.50}{(8{,}044.931 \times 0.50) + 0.50} = 0.999987$$

The genetic evidence (CPI) has changed the probability of paternity (prior odds) of 50% to 99.9987%.

There is some disagreement about the assumption of 50% prior odds. Using different prior odds assumptions changes the final probability of paternity (Table 10.6). As can be observed from the table, however, at a CPI over 100, the differences have less effect.

Sibling Tests

Polymorphisms are also used to generate a probability of siblings or other blood relationships (familial searches).[24] A sibling test is a more complicated statistical analysis than a paternity test. Mutations and allele frequencies further complicate analysis.[25] More confident conclusions can be made with multiple siblings. Methods involving parental genotype reconstruction have been proposed.[26] A full sibling test is a determination of the likelihood that two people tested share a common mother and father. A half-sibling test is a determination of the likelihood that two people tested share one common parent (mother or father). The likelihood ratio generated by a sibling test is sometimes called a **kinship index, sibling index,** or **combined sibling index.**

Another type of relationship analysis is **avuncular testing,** which measures the probabilities that two alleged relatives are related as either an aunt or an uncle of a niece or nephew. The probability of relatedness is based on the number of shared alleles between the tested individuals. As with paternity and identity testing, allele frequency in the population will affect the significance of the final results. The probabilities can be increased greatly if other known relatives, such as a parent of the niece or nephew, are available for testing. Determination of first- and second-degree relationships is important for genetic studies because linkage mapping of disease genes in populations can be affected by undetected familial relationships.[27]

Y-STR

Unlike conventional STRs **(autosomal STRs),** where each locus is defined by two alleles, one from each parent, Y-STRs are represented only once per genome and only in males (Fig. 10.15). A set of Y-STR alleles comprises a **haplotype,** or series of linked alleles always inherited together. This is because the Y chromosome cannot extensively exchange information (recombine) with the X chromosome or another Y chromosome. Thus, marker alleles on the Y chromosome are inherited from generation to generation in a single block. This means that the frequency of entire Y-STR profiles (haplotypes) in a given population can be determined by empirical studies. For example, if a combination of alleles (haplotype) was observed only two times in a test of 200 unrelated males, that haplotype is expected to occur with a frequency of approximately 1 in 100 males tested in the future. The discrimination power of Y-haplotype testing will depend on the number of

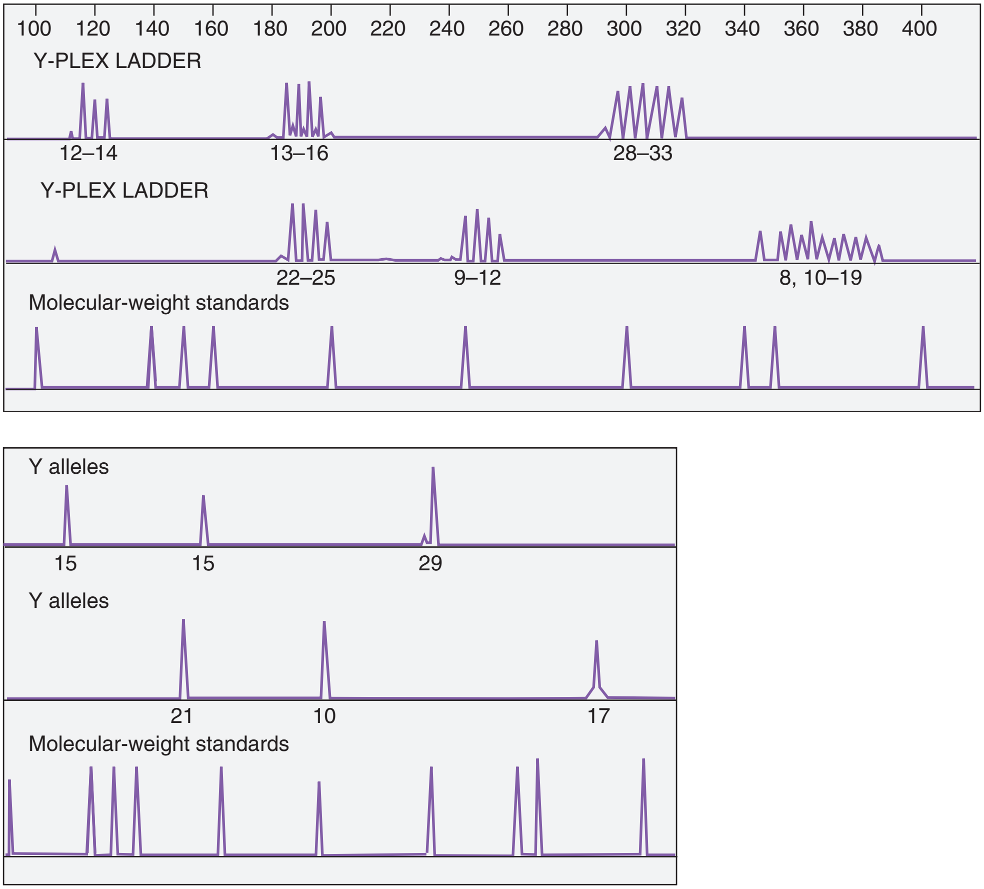


FIGURE 10.15 Electropherogram showing allelic ladders for six STR loci in the Y-Plex 6 system *(top panel)* and a single haplotype *(bottom panel)*. Molecular-weight standards are shown at the bottom of each.

subjects tested and will be less definitive than that of autosomal STR.

Despite being a less powerful system for identification, STR polymorphisms on the Y chromosome have unique characteristics that have been exploited for forensic, lineage, and population studies as well as kinship testing.[28] Except for rare mutation events, every male member of a family (brothers, uncles, cousins, and grandfathers) will have the same Y-chromosome haplotype. Thus, Y-chromosome inheritance can be applied to lineage, population, and human migration studies. The **Y-STR/paternal lineage test** can determine whether two or more males have a common paternal ancestor. In addition to family history studies, the results of a paternal lineage test serve as supportive evidence for adoptees and their biological relatives or for individuals filing inheritance and benefit claims. Because Y chromosomes are inherited intact, spontaneous mutations in the DNA sequence of the Y chromosome are used to follow human migration patterns and historical lineages. Y-chromosome genotyping has been used to locate the geographical origin of populations.

Because all male relatives in a family will share the same allele combination or profile, the statistical significance of a Y-STR DNA match cannot be assessed by multiplying likelihood ratios as was described previously for autosomal STR. Instead of the allele frequency used in autosomal STR match calculations, haplotype frequencies are used. Estimation of haplotype frequencies, however, is limited by the number of known Y haplotypes. This smaller data set accounts for the reduced inclusion probabilities and a discrimination rate that is significantly lower than that for autosomal STR polymorphisms. Traditional STR loci are, therefore, preferred for identity or relationship analyses, and the Y-STRs are used to aid in special situations—for instance, in confirming sibship between males who share commonly occurring alleles, that is, have a low likelihood ratio based on traditional STRs.

Y-STRs have been utilized in forensic tests where the evidence consists of a mixture of male and female DNA, such as semen, saliva, other body secretions, or fingernail scrapings. For instance, in specimens from the evidence of sexual assault, the female DNA may be in vast excess (more than 100-fold) compared to the male DNA in the sample. Autosomal STRs are not consistently informative under these circumstances. Using Y-specific primers, Y-STR can be specifically amplified by PCR from the male–female mixture, resulting in an analyzable marker that has no female background. This affords a more accurate identification of the male donor.

The Y chromosome has a low mutation rate. The overall mutation rate for Y chromosome loci is estimated at 7.4×10^{-10} mutations per position per year.[29] Assuming that Y-chromosome mutations generally occur once every 500 generations/locus, for 25 loci, 1 locus should have a mutation every 20 generations (500 generations/25 markers = 20 generations). Lineage testing over several generations is made possible by this low mutation rate. It is also useful for missing persons' cases in which reference samples can be obtained from paternally related males.

A list of informative Y-STRs is shown in Table 10.7. Several Y-STRs are located in regions that are duplicated on the Y chromosome. DSY389I and DSY389II are examples of duplicated loci.

Like autosomal STRs, Y-STRs have microvariant alleles containing incomplete repeats and alleles containing repeat sequence differences. Reagent systems consisting of multiplexed primers for identification of 6-17 Y-STR loci are available commercially.

Matching With Y-STRs

Matching probabilities from Y-STR data are determined differently than for the autosomal STR. **Haplotype diversity (HD)** is calculated from the frequency of occurrence of a given haplotype in a tested population. The probability of two random males sharing the same haplotype is estimated at 1-HD; that is, if the haplotype diversity is high, the probability of two random males in the population having the same haplotype is low. Another measure of profile uniqueness, the **discriminatory capacity (DC),** is determined by the number of different haplotypes seen in the tested population and the total number of samples in the population. DC expresses the percentage of males in a population who can be identified by a given haplotype. Just as the number of loci included in an autosomal STR genotype increases the power of discrimination, DC is increased by increasing the number of loci defining a haplotype. For instance, six loci tested can distinguish 82% of African American males. Using 22 loci raises the DC to almost 99% (Table 10.8).

TABLE 10.7 Y-STR Locus Information*[72-74]

Y-STR	Repeat Sequence†	Alleles
DYS19	[TAGA]^{3}TAGG[TAGA]$_{n}$	10, 11, 12, 13, 14, 15, 16, 17, 18, 19
DYS385	[GAAA]$_{n}$	7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 16.3, 17, 17.2 17.3, 18, 19, 20, 21, 22, 23, 24, 28
DYS388	[CAA]$_{n}$	10, 11, 12, 13, 14, 15, 16, 17, 18
DYS389 I‡	[TCTG]^{q} [TCTA]^{r}	9, 10, 11, 12, 13, 14, 15, 16, 17
DYS389 II‡	[TCTG]$_{n}$[TCTA]^{p}[TCTG]^{q}[TCTA]^{r}	26, 27, 28, 28′, 29, 29′, 29″, 29‴, 30, 30′ 30″, 30‴, 31, 31′, 31″, 32, 32′, 33, 34
DYS390	[TCTG]$_{n}$[TCTA]^{m}[TCTG]^{p}[TCTA]	17,18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28
DYS391	[TCTA]$_{n}$	6, 7, 8, 9, 10, 11, 12, 13, 14
DYS392	[TAT]$_{n}$	6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17
DYS393	[AGAT]$_{n}$	9, 10, 11, 12, 13, 14
DYS426	[CAA]$_{n}$	6.2, 9, 10, 11, 12, 13, 14
DYS434	[CTAT]$_{n}$	8, 9, 10, 11
DYS437	[TCTA]$_{n}$[TCTG]^{2}[TCTA]^{4}	13, 14, 15, 16, 17
DYS438		
DYS439	[TTTTC]$_{n}$	6, 7, 8, 9, 10, 11, 12, 13, 14
DYS439	[GATA]$_{n}$	9, 10, 11, 12, 13, 14
(Y-GATA-A4)	[GATA]$_{n}$	9, 10, 11, 12, 13, 14
DYS441	[CCTT]$_{n}$	8, 10.1, 11, 11.1, 12, 13, 13.1, 14, 14.3, 15, 16, 17, 18, 19, 20
DYS442	[TATC]$_{n}$	8, 9, 10, 11, 12, 12.1, 13, 14, 15
DYS444	[TAGA]$_{n}$	9, 10, 11, 12, 13, 14, 15, 16
DYS445	[TTTA]$_{n}$	6, 7, 8, 9, 10, 10.1, 11, 12, 13, 14
DYS446	[AGAGA]$_{n}$	8, 9, 10, 11, 12, 13, 14, 15, 15.1, 16, 17, 18, 19, 19.1, 20, 21, 22, 23
DYS447	[TTATA]$_{n}$	15, 16, 17, 18, 19, 19.1, 20, 21, 22, 22.2, 22.4, 23, 24, 25, 26, 26.2, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36
DYS448	[AGAGAT]$_{n}$	17, 19, 19.2, 20, 20.2, 20.4, 21, 21.2, 21.4, 22, 22.2, 23, 23.4, 24, 24.5, 25, 26, 27
DYS449	[GAAA]$_{n}$	23, 23.4, 24, 24.5, 25, 26, 27, 27.2, 28, 28.2, 29, 29.2, 30, 30.2, 31, 32, 32.2, 33, 33.2, 34, 35, 36, 37, 37.3, 38
DYS452	[TATAC]$_{n}$	24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35
DYS454	[TTAT]$_{n}$	6, 7, 8, 9, 10, 11, 12, 13
DYS455	[TTAT]$_{n}$	7, 8, 9, 10, 11, 12, 13

TABLE 1.7 Y-STR Locus Information *(Continued)*

Y-STR	Repeat Sequence[†]	Alleles
DYS456	$[AGAT]_n$	11, 12, 13, 14, 15, 16, 17, 18, 19, 20
DYS458	$[CTTT]_n$	12, 12.2, 13, 14, 15, 15.2, 16, 16.1, 16.2, 17, 17.2, 18, 18.2, 19, 19.2, 20, 20.2, 21
DYS460 (Y-GATA-A7.1)	$[ATAG]_n$	7, 8, 9, 10, 10.1, 11, 12, 13

*http://www.cstl.nist.gov

[†] Some alleles contain repeats with one, two, or three bases missing.

[‡] DYS389 I and II is a duplicated locus.

TABLE 10.8 Discriminatory Capacity of Y-STR Genotypes in Different Subpopulations[75]

	African American (%)	White American (%)	Hispanic American (%)
6 loci*	82.3	68.9	78.3
9 loci[†]	84.6	74.8	85.1
11 loci[‡]	91.3	83.8	90.3
17 loci	99.1	98.8	98.3
20 loci[§]	98.5	97.2	98.6
22 loci[‖]	98.9	99.6	99.3

*Y-Plex 6 (DYS19, DYS390, DYS391, DYS393, DYS389II, and DYS385).

[†] European minimal haplotype.

[‡] Minimal haplotype + SWGDAM.

[§] Y-STR 20 plex (minimal haplotype plus DYS388, DYS426, DYS437, DYS439, DYS460, H4, DYS438 DYS447, and DYS448).

[‖] Y-STR 22 plex.

Advanced Concepts

The European Y chromosome typing community has established a set of Y-STR loci termed the **minimal haplotype** (see http://www.yhrd.org). The minimal haplotype consists of Y-STR markers DYS19, DYS389I, DYS389II, DYS390I, DYS391, DYS392, DYS393, and DYS385.[30] An **"extended haplotype"** includes all of the loci from the minimal haplotype plus the highly polymorphic dinucleotide repeat YCAII.

Because there is no recombination between loci on the Y chromosome, the product rule cannot be applied. The results of a Y typing might be reported accompanied by the number of observations or frequency of the analyzed haplotype in a database of adequate size. Suppose a haplotype containing the 17 allele of DYS390 occurs in only 23% of men in a database of 12,400. However, if that same haplotype contains the 21 allele of DYS446, only 6% of the men will have haplotypes containing both the DYS390 17 and DYS446 21 alleles. If the 11 allele of DYS455 and the 15 allele of DYS458 are also present, only 1 out of 12,400 men in the population has a haplotype containing all four alleles. The uniqueness of this haplotype is strong evidence that a match is not the result of a random coincidence, which gives extra support to the hypothesis that an independent source with this haplotype comes from an individual or a paternal relative. Even with a 1/12,400 or 99.9% DC, however, the matching probability is orders of magnitude lower than that for autosomal STR.

Y-chromosome haplotypes can be used to exclude paternity. Taking into account the mutation rate of each allele, any alleles that differ between the male child and the alleged father are strong evidence for non-paternity. Conversely, if a Y haplotype is shared between a child and alleged father, a paternity index is calculated in a manner similar to that of the autosomal STR analysis. For example, suppose 6 Y-STR alleles are tested and match between the alleged father and child. If the haplotype has not been observed before in the population, the occurrence of that haplotype in the population database is 0/1,200, and the haplotype frequency will be 1/1,200, or 0.0008333. The paternity index (PI) is the probability that an alleged father with that haplotype could produce one sperm carrying the haplotype, divided by the probability that a random man could produce one sperm carrying the haplotype. The PI is then 1/0.0008333 = 1,200. With a prior probability of 0.5, the probability of paternity is

$$(1,200\times 0.5)/[(1,200\times 0.5)+0.5], \text{ or } 99.9\%.$$

This result, however, does not exclude patrilineal relatives of the alleged father.

Y-STRs as marker loci for Y-chromosome, or **surname,** tests are used to determine ancestry. For example, a group of males of a strictly male descent line (having the same last name or surname) is expected to be related to a common male ancestor. Therefore, they should all share the same Y-chromosome alleles (except for mutations, which should be minimal, given 1 mutation per 20 generations, as explained previously). The Y-chromosome haplotype does not provide information about the degree of relatedness, just inclusion or exclusion from a family. An analysis to find a most recent common ancestor (MRCA) is possible, however, using a combination of researched family histories, Y-STR test results, and statistical formulas for mutation frequencies.

LINKAGE ANALYSIS

Because the locations of many STRs in the genome are known, these structures can be used to map genes, especially those genes associated with disease. Three basic approaches are used to map genes: family histories, population studies, and sibling analyses.

Family history and analysis of generations of a single family for the presence of a particular STR allele in affected individuals is one way to show association. Family members are tested for several STRs, and the alleles of affected and unaffected members of the family are compared. Assuming normal Mendelian inheritance, if a particular allele of a particular locus is always present in affected family members, that locus must be closely linked to the gene responsible for the phenotype in those individuals (**linkage disequilibrium;** Fig. 10.16). If the linkage is close enough to the gene (no recombination between the STR and the disease gene), the STR may serve as a convenient marker for disease testing. Instead of testing for mutations in the disease gene, the marker allele is determined. It is easier, for example, to look for a linked STR allele than to screen a large gene for point mutations. The presence of the "indicator" STR allele serves as a genetic marker for the disease (Fig. 10.17).

Another approach to **linkage analysis** is to look for gene associations in large numbers of unrelated individuals in population studies. Just as with family history studies, close linkage to specific STR alleles supports the genetic proximity of the disease gene with the STR. In this case, however, large numbers of unrelated people are tested for linkage rather than a limited number of related individuals in a family. The results are expressed in probability terms that an individual with the linked STR allele is likely to have the disease gene. With the accumulation of genomic data produced by massive parallel sequencing methods, however, this type of population study is currently done with higher resolution using

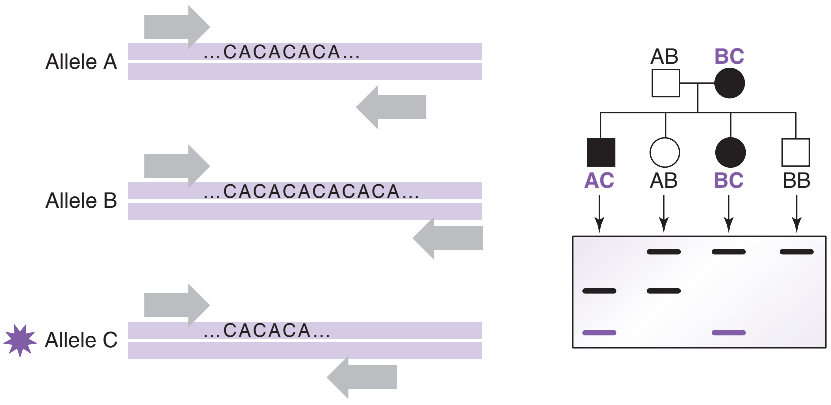


FIGURE 10.16 Linkage analysis with STRs. Three alleles, A, B, and C, of an STR locus are shown *(left)*. At right is a family pedigree showing assortment of the alleles along with gel analysis of PCR amplification products. Allele C is present in all affected family members. This supports the linkage of this STR with the gene responsible for the disease affecting the family. Analysis for the presence of allele C of this STR may also provide a simple indicator to predict inheritance of the affected gene.

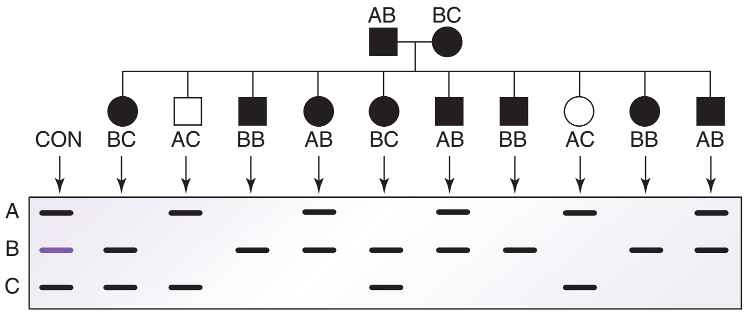


FIGURE 10.17 Inheritance of alleles in an affected family. Using the banding pattern shown, the B allele of this STR is always present in affected individuals. This locus must be closely linked to the mutated gene.

SNPs, rather than STR (see the following discussion). Either type of marker is informative if a linkage is found.

Sibling studies are another type of linkage analysis. Monozygotic (identical) and dizygotic (fraternal) twins serve as controls for genetic and environmental studies. Monozygotic twins will always have the same genetic alleles, including disease genes. There should be 100% recurrence risk (likelihood) that if one twin has a genetic disease, the other twin has it, and both should have the same linked STR alleles. Fraternal twins have the same likelihood of sharing a gene allele as any sibling pair. Investigation of adoptive families may also distinguish genetic from environmental or somatic effects.

BONE MARROW ENGRAFTMENT TESTING USING DNA POLYMORPHISMS

Bone marrow transplantation is a method used to treat malignant and nonmalignant blood disorders as well as some solid tumors. One transplant approach is **autologous** (from self), in which cells from the patient's own bone marrow are removed and stored. The patient then receives high doses of chemotherapy and/or radiotherapy. The portion of marrow previously removed from the patient may also be purged of cancer cells before being returned to the patient. Alternatively, **allogeneic** transplants (between two individuals) are used. The donor supplies healthy cells to the recipient patient (Fig. 10.18). Donor cells are supplied as bone marrow, peripheral blood stem cells (also called **hematopoietic stem cells**), or umbilical cord stem cells. To assure successful establishment of the transplanted donor cells, the immune compatibility of the donor and recipient is tested prior to the transplant by HLA typing.

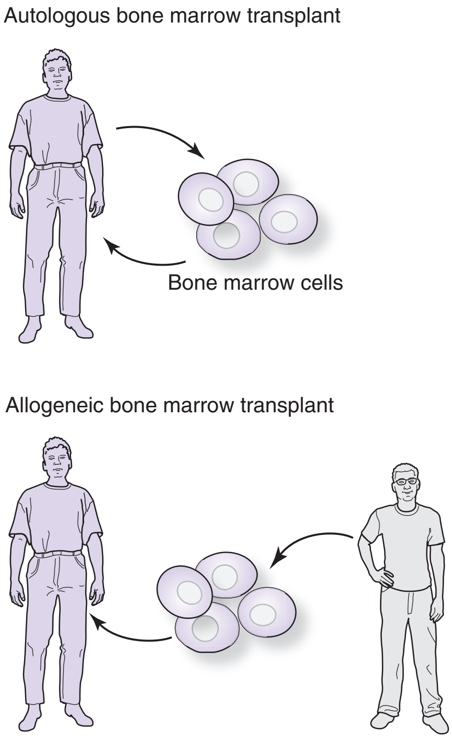

Autologous bone marrow transplant
Bone marrow cells
Allogeneic bone marrow transplant

FIGURE 10.18 In autologous bone marrow transplant *(top)*, bone marrow cells are taken from the patient, purged, and returned to the patient after conditioning treatment. In allogeneic transplant *(bottom)*, bone marrow cells are taken from another genetically compatible individual (donor, *right*) and given to the patient *(left)*.

Historical Highlights

The first successful bone marrow transplants were syngeneic transplants; that is, the donor and recipient pairs were identical twins. In the late 1950s, compatibility between donor and recipient HLAs was found to be necessary for the success of allogeneic transplants. With the development of genetically defined animals, cross-species or xenogeneic transplants were considered, and in 1992, a xenotransplant from baboon to human was attempted. The patient died 26 days later from infection. Currently, donor registries and advances in the use of hematopoietic stem cells have broadened the application of transplants for a variety of diseases.

In allogeneic transplant strategies, high doses of therapy completely remove the recipient bone marrow, particularly the stem cells that give rise to all the other cells in the marrow (conditioning). The allogeneic stem cells are then expected to reestablish a new bone marrow in the recipient **(engraftment).** The toxicity of this procedure was reduced by using sub-myeloablative transplant procedures or mini-transplants. In this approach, pretransplant therapy will not completely remove the recipient bone marrow. The donor bone marrow is expected to eradicate the remaining recipient cells through recognition of residual recipient cells as foreign to the new bone marrow. This process also imparts a graft-versus-leukemia (GVL) or **graft-versus-tumor (GVT) effect,** which is a process closely related to **graft-versus-host disease (GVHD).** The T-cell fraction of the donor marrow is particularly important for engraftment and for GVT effect. This was realized when efforts to avoid GVHD by removing the T-cell fraction before infusion of donor cells resulted in increased incidence of graft failure and relapse.

The first phase of allogeneic transplantation is donor matching, in which potential donors are tested for immunological compatibility. This is performed by examining the HLA locus to determine which donor would be most tolerated by the recipient immune system. Donors may be known or related to the patient or anonymous unrelated contributors **(matched unrelated donor [MUD]).** The National Marrow Donor Program (NMDP) maintains a database of people who have voluntarily submitted their HLA types and are willing to serve as potential donors.

Stem cells may also be acquired from donated umbilical cord blood. After conditioning and infusion with the donor cells, the patient enters the engraftment phase, in which the donor cells reconstitute the recipient's bone marrow. Once a successful engraftment of donor cells is established, the recipient is a genetic **chimera;** that is, the recipient has tissue and blood cells of separate genetic origins.

The engraftment of donor cells in the recipient must be monitored, especially in the first 90 days after the transplant. This requires distinguishing between donor cells and recipient cells. Historically, red blood cell phenotyping, immunoglobulin allotyping, HLA typing, karyotyping, and fluorescence in situ hybridization have been used for this purpose. Each of these methods has drawbacks. Some require months before engraftment can be detected. Others are labor-intensive or restricted to sex-mismatched donor–recipient pairs.

DNA typing has become the method of choice for engraftment monitoring.[31,32] Because all individuals, except identical twins, have unique DNA polymorphisms, donor cells are monitored by following donor polymorphisms in the recipient blood and bone marrow. Although RFLP can effectively distinguish donor and recipient cells, the detection of RFLP requires the use of the Southern blot method, which has limited sensitivity for this application. In comparison, small VNTRs and STRs are easily detected by PCR amplification of VNTRs and STRs (see Fig. 10.9), which is preferable because of the increased rapidity and the 0.5% to 1% sensitivity achievable with PCR. Sensitivity can be raised to 0.01% using Y-STR, but this approach is limited to those transplants from sex-mismatched donor–recipient pairs, preferably from a female donor to a male recipient.

Advanced Concepts

Chimerism is different from mosaicism. A chimera is an individual carrying two populations of cells that arose from different zygotes. In a **mosaic,** cells arising from the same zygote have undergone a genetic event, resulting in two clones of phenotypically different cells in the same individual.

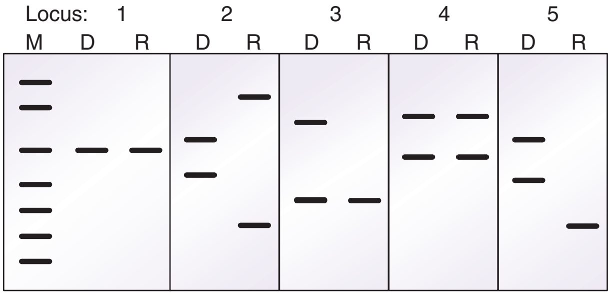


FIGURE 10.19 Band patterns of five different loci comparing donor (D) and recipient (R) genotypes. The second and fifth loci are informative. The first and fourth loci are noninformative. The third locus is donor-informative.

In the laboratory, there are two parts to engraftment/chimerism DNA testing. Before the transplant, several polymorphic loci in the donor and recipient cells are screened to find at least one **informative locus,** that is, one locus in which donor alleles differ from the recipient alleles. **Noninformative loci** are those in which the donor and the recipient have the same alleles. In donor-informative loci, the donor and recipient share one allele for which the donor is heterozygous and the donor has a unique allele. Conversely, in recipient-informative loci, the unique allele is in the recipient (Fig. 10.19). The second part of the testing process is the engraftment analysis, which is performed at specified intervals after the transplant. In the engraftment analysis, the recipient blood and bone marrow are tested to determine the presence of donor cells using the informative and/or recipient-informative loci.

Pretransplant analysis and engraftment were measured in early molecular studies by amplification of small VNTRs and resolution of amplified fragments on polyacrylamide gels with silver-stain detection.[33] Before the transplant, the screen for informative loci was based on band patterns of the PCR products, as illustrated in Figure 10.19. After the transplant, analysis of the gel band pattern from the blood or bone marrow of the recipient revealed one of three different states: **full chimerism,** in which only the donor alleles were detected in the recipient; **mixed chimerism,** in which a mixture of donor and recipient alleles was present; or **graft failure,** in which only recipient alleles were detectable (Fig. 10.20).

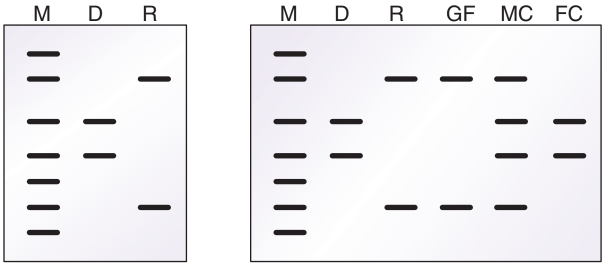


FIGURE 10.20 Band patterns after PAGE analysis of VNTR. First, before the transplant, several VNTR must be screened to find informative loci that differ in pattern between the donor and recipient. One such marker is shown at left (M = molecular-weight marker, D = donor, R = recipient). After the transplant, the band patterns can be used to distinguish between graft failure (GF), mixed chimerism (MC), or full chimerism (FC).

Currently, the preferred method is PCR amplification of STRs, resolution by capillary electrophoresis, and fluorescent detection. This procedure provides ease of use, accurate quantification of the percentage of donor/recipient cells, and high sensitivity with minimal sample requirements. An alternate, even more sensitive method using qPCR of SNP has been proposed.[34] This method also has the advantage of higher throughput and lower sample requirements. It can be performed on a 96-well plate format as a sequence-specific qPCR with no gel resolution required. Limited informative SNP and the requirement to include pretransplant donor and recipient DNA at each monitoring time point have slowed the adoption of this method in many laboratories.

PSTR Testing

Donor and recipient DNA for allele screening prior to transplant can be isolated from blood or buccal cells. The lower limit of 1 ng of DNA is reportedly sufficient for the screening of multiple loci; however, 10 ng is a more practical lower limit. Multiple loci can be screened simultaneously using multiplex PCR. Although not validated for engraftment testing, several systems designed for human identification are used for this purpose. Primer sets that specifically amplify Y-STR may also be useful for sex-mismatched donor–recipient pairs. Figure 10.21 shows the five tetramethylrhodamine (TMR)-labeled loci originally developed for identity testing. Although multiplex primer systems are optimized for consistent results, all loci may not amplify with equal efficiency in a multiplex reaction. For example, the amelogenin locus in Figure 10.21 did not amplify as well as the other four loci in the multiplex. This is apparent from the lower peak heights in the amelogenin products compared with the products of the other primers. Less efficient amplification lowers the sensitivity of subsequent post-engraftment testing.

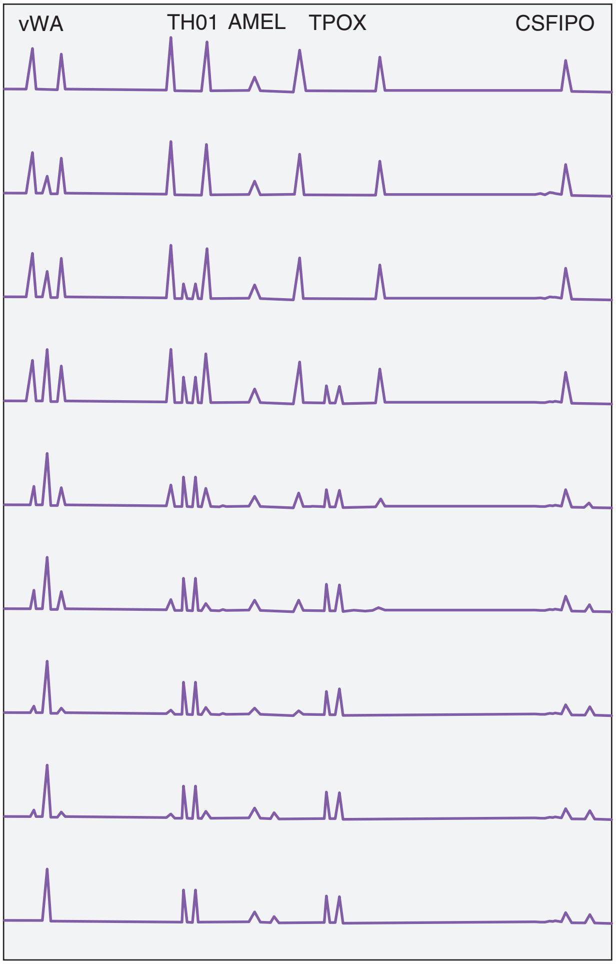


FIGURE 10.21 Multiplex PCR of DNA mixtures from two unrelated individuals (top and bottom trace) showing peak patterns for vWA, TH01, Amelogenin, TPOX, and CSF1PO loci. The center traces are stepwise mixtures of the two genotypes.

Advanced Concepts

A more defined condition can be uncovered by cell type separation. Some cell fractions such as granulocytes engraft before others. For example, isolated granulocytes may show full chimerism, whereas the T-cell fraction still shows mixed chimerism. This is a case of **split chimerism.**

Although the capillary electrophoresis used for this method is the same as that used for sequence analysis, measuring peak sizes and peak areas is distinguished from sequence analysis as **fragment analysis** and sometimes requires adjustment of the instrument or capillary polymer. Automatic detection will generate an electropherogram, as shown in Figure 10.22. Informative and noninformative loci will appear as nonmatching or matching donor and recipient peaks, respectively, and many combinations of donor–recipient peaks are possible. Optimal loci for analysis should be clean peaks without stutter, especially stutter peaks that co-migrate

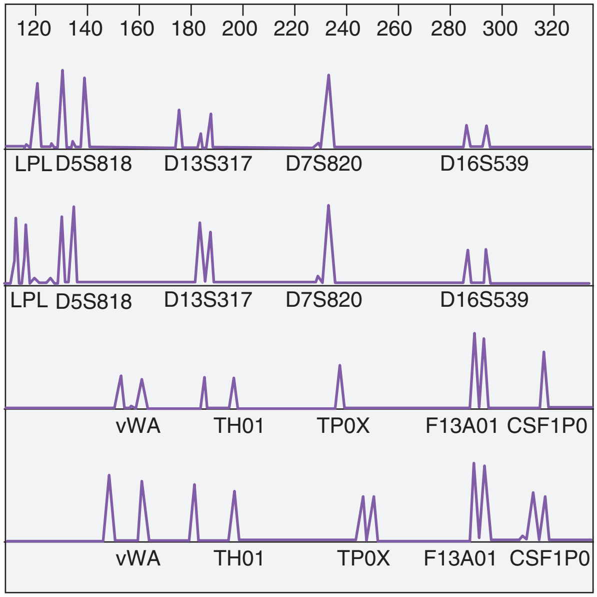


FIGURE 10.22 Screening of 16 loci for informative STR alleles. Recipient peak patterns (first and third traces) are compared with donor patterns (second and fourth). LPL, D5S818, D13S317.VWA, THO1, and TPOX are informative. CSF1PO is donor-informative.

with informative peaks, nonspecific amplified peaks (mis-primes), or other technical artifacts.[35]

Ideally, the chosen locus should have at least one recipient informative allele. This is to assure direct detection of minimal amounts of residual recipient cells. If the recipient is male and the donor is female, the amelogenin locus supplies a recipient-informative locus. Good separation (ideally, but not necessarily, by two repeat units) of the recipient and donor alleles is desirable for ease of discrimination in the post-transplant testing. The choices of informative alleles are more limited in related donor–recipient pairs, as they are likely to share alleles. Unrelated donor–recipient pairs, on the other hand, will yield more options.

After the transplant, the recipient is tested on a schedule determined by the clinician or according to consensus recommendations.[36] The frequency of testing should be determined by clinical standard operating procedures. Testing will depend on the disease and individual patient assessment. With nonmyeloablative or reduced-intensity pretransplant protocols, an example schedule would be testing at 1, 3, 6, and 12 months. Because early patterns of engraftment may predict GVHD or graft failure after non-myeloablative treatments, even more frequent blood testing may be necessary, such as at 1, 2, and 3 months after transplant. Bone marrow specimens can most conveniently be taken at the time of bone marrow biopsy following the transplant, with blood specimens taken in intervening periods. Usually, 3 to 5 mL of bone marrow or 5 mL of blood is more than sufficient for analysis; however, specimens collected soon after the transplant may be hypocellular, so larger volumes (5 to 7 mL bone marrow, 10 to 20 mL blood) may be required.

Post-Transplant Engraftment Testing

Quantification of the percentage of recipient and donor cells post-transplant is performed using the informative locus or loci selected during the pretransplant informative analysis. The raw data for these calculations are the peak heights or areas under the peaks generated by the PCR products after amplification. Peaks are generated by the emission from the fluorescent dyes attached to the primers and thus to the ends of the PCR products collected as each product migrates past the detector. The fluorescent signal is converted into fluorescence units by the computer software. The software displays the PCR products as peaks of fluorescence units (y-axis) versus migration speed (x-axis). The amount of fluorescence in each product or peak, represented as the height or area under the peak, is used to calculate the percentage of recipient and donor cells (Fig. 10.23).

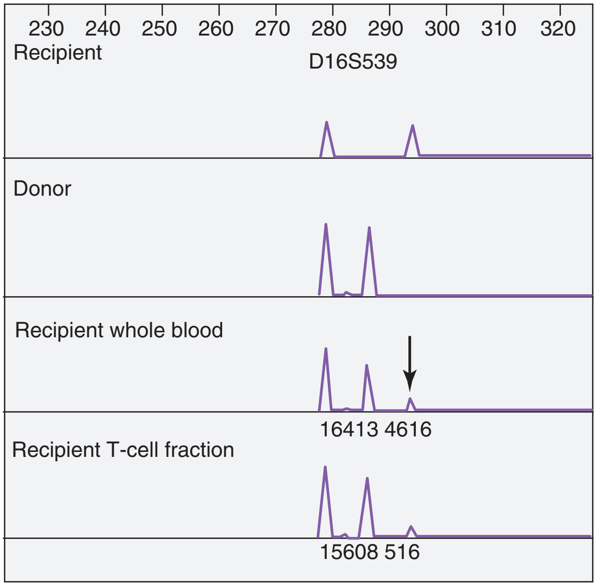


FIGURE 10.23 Post-engraftment analysis of an informative locus D16S539. The area (fluorescence units) under the peaks is calculated automatically. The recipient and donor patterns are shown in the first and second trace, respectively. Results from the whole blood and T-cell fraction are shown in the third and fourth traces. For D16S539, the formula $R_{(unshared)}/(R_{(unshared)} + D_{(unshared)})$ yields 4,616/(4,616 + 16,413) × 100 = 22% recipient cells in the unfractionated blood (arrow) and 516/(516 + 15,608) × 100 = 3.2% recipient cells in the T-cell fraction.

Advanced Concepts

Occasionally, specimens may be received in the laboratory after engraftment without pre-engraftment information. In this case, the blood or bone marrow of the recipient is not acceptable for determination of recipient-specific alleles because the alleles present may represent donor, recipient, or both donor and recipient. The specimen can be processed using the amelogenin locus or Y-STR markers if the donor and recipient are of different genders, preferably a female donor and a male recipient. Another option is to use an alternate source of recipient DNA, such as buccal cells, skin biopsy sample, or stored specimens or DNA from previous testing. Because of the nature of lymphocyte migration, however, skin and buccal cells may also have donor alleles due to the presence of donor lymphocytes in these tissues. The best approach is to ensure that informative analysis of the donor and recipient are part of the pretransplant agenda.

There are several formulae for percent calculations, depending on the configuration of the donor and recipient peaks. For homozygous or heterozygous donor and recipient peaks with no shared alleles, the percentage of recipient cells is equal to R/(R + D), where R is the height or area under the recipient-specific peak(s), and D is the height or area under the donor-specific peak(s). Shared alleles, where one allele is the same for donor and recipient (Fig. 10.22), can be dropped from the calculation, and the percentage of recipient cells is calculated as

$$\frac{R_{(\text{unshared})}}{(R_{(\text{unshared})} + D_{(\text{unshared})})}$$

Chimerism/engraftment results are reported as the percentage of recipient cells and/or percentage of donor cells in the bone marrow, blood, or cell fraction. These results do not reflect the absolute cell number, which could change independently of the donor/recipient ratio. Inability to detect donor or recipient cells does not mean that the cell population is completely absent because capillary electrophoresis and fluorescent detection methods offer a sensitivity of 0.1% to 1% for autosomal STR markers. Time trends may be more important than single-point results following transplantation.

Advanced Concepts

Positive or negative selection techniques may be used to test specific cell lineages. For example, analysis of the T-cell fraction separately is used to monitor graft-versus-tumor potential. T cells may comprise 10% of peripheral blood leukocytes and 3% of bone marrow cells following allogeneic transplantation. Analysis of unfractionated blood and especially bone marrow where all other lineages are 100% could miss split chimerism in the T-cell fraction. T-cell-lineage-specific chimerism will therefore increase the sensitivity of the engraftment analysis, particularly after nonmyeloablative and immunoablative pretransplant treatments.

Testing of the engraftment of lineage-specific call populations increases the informativity and sensitivity of the engraftment testing. Selected cell populations (e.g., T or B lymphocytes, or myeloid cells) are conveniently separated from whole blood using magnetized polymer particles (beads) attached to cell type–specific antibodies, such as pan-T antibodies (anti-CD3) to isolate T cells. To isolate cells using beads, whole blood or bone marrow or white blood cells isolated by density-gradient centrifugation are mixed with the beads in saline or phosphate-buffered saline and incubated to allow the antibodies on the beads to bind to the antigens on the cell surface. With the beads, cells with bound beads are immobilized by a magnet that is applied to the outside of the tube, and the supernatant containing cells without bound beads is decanted. After another saline wash, the cells are collected and lysed for DNA isolation. It is not necessary to detach the cells from the beads for most procedures.

Automated cell sorter systems may also be used for this purpose. With a positive selection program, for example, the instrument is capable of isolating up to 2×10^8 pure T cells per separation. Unwanted cells can be removed with the depletion programs.

Because cell lineages engraft with different kinetics, testing of blood and bone marrow may yield different levels of chimerism. Bone marrow will contain more myeloid cells, and blood will contain more lymphoid cells. The first determination to be made from engraftment testing is whether donor engraftment has occurred and, secondly, whether there is split chimerism. In split chimerism, cell separation techniques may be used to determine which lineages are mixed and which are fully donor.

QUALITY ASSURANCE FOR SURGICAL SECTIONS USING STR

The molecular diagnostics laboratory can assist in ensuring that surgical tissue sections are properly identified and not contaminated. During processing of tissue specimens, microscopic fragments of tissue may persist in paraffin baths (floaters).[37] These fragments can adhere to subsequent tissue sections, resulting in the anomalous appearance of the tissue under the microscope. If a tissue sample is questioned, STR identification can confirm the origin of tissue.[38,39,45,46]

For this procedure, the suspect tissue must be carefully removed from the slide by microdissection. Reference DNA isolated from the patient and DNA isolated from the tissue in question are subjected to multiplex PCR. The results are compared for matching alleles. If the tissue in question originated from the patient, all alleles should match. Assuming good-quality data, one nonmatching locus excludes the tissue in question as coming from the reference patient.

An example of such a case is shown in Figure 10.24. A uterine polyp was removed from a patient for microscopic examination. An area of malignant tissue was present on the slide. The pathologists were suspicious about the malignancy because there was no other malignant tissue observed in other tissue from the patient. The tissue fragment was microdissected from the thin section and tested at nine STR loci. The allelic profile was compared to reference DNA from the patient. The profiles were identical, confirming that the tissue fragment was from the patient.

Using STR for this application has some limitations. Tissue quality and quantity can adversely affect amplification of the STR loci, especially the larger products. Also, DNA isolated from small fragments may be mixed with reference DNA, complicating interpretation and comparison of alleles. A reported case study also demonstrated the effect of inherent genomic instability in tumor tissue.[40] Allele differences between the suspected floater of malignant tissue and reference tissue from the patient led to the initial conclusion that it was from another source. A second biopsy, however, also contained malignant cells similar to those found in the first biopsy. The allele differences were determined to be a result of microsatellite instability in the malignant cells. Another study comparing STR alleles in genetically stable and unstable tumors directly demonstrated the presence of new alleles in unstable tumor tissue compared to normal reference tissue from the same patient.[41] It is advisable, therefore, to take into account whether tumor cells might be genetically unstable when testing for contaminants.

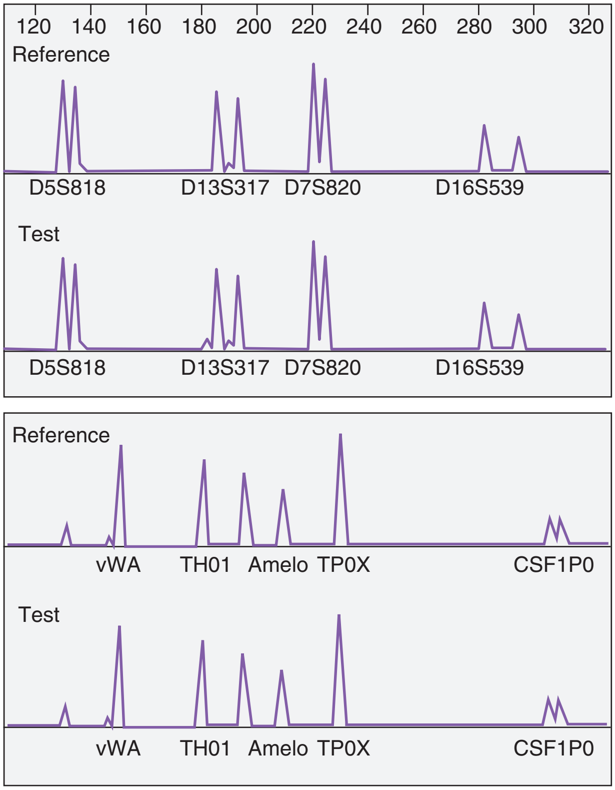


FIGURE 10.24 Quality assurance testing of a tissue fragment. The STR profile of the fragment in question (test) was compared with that of reference DNA from the patient. The alleles matched at all loci, supporting the genotypic identity of the test material with the patient.

SINGLE-NUCLEOTIDE POLYMORPHISMS

Data from the Human Genome Project revealed that the human nucleotide sequence differs every 1,000 to 1,500 bases from one individual to another.[42] The majority of these sequence differences are variations of single nucleotides, or SNPs. The traditional definition of polymorphism requires that the genetic variation be present at a frequency of at least 1% of the population. The International SNP Map Working Group observed that two haploid genomes differ at 1 nucleotide per 1,331 bp.[43] This rate, along with the theory of neutral changes expected in the human population, predicts 11 million sites in a genome of 3 billion bp that vary in at least 1% of the world's population. In other words, each individual has 11 million SNPs. Higher-density analysis made possible by next-generation sequencing has determined that SNPs are even more frequent—one SNP per every 300 nucleotides in a given human genome.[44]

Due to the density of SNPs across the human genome, these polymorphisms were of great interest for genetic mapping, disease prediction, and human identification. The problem was that detection of single-base-pair changes was not as easy as the detection of STRs, VNTRs, or even RFLPs.

With improving technology, mapping studies achieved denser coverage of the genome.[45] The most definitive way to detect SNPs has been by direct sequencing. A number of additional methods were subsequently designed to detect known SNPs. Next-generation sequencing has greatly accelerated both the discovery and detection of SNPs.[46,47] Computer analysis is also required to confirm that the population frequency of the SNPs meets the requirements of a polymorphism. So far, approximately 10 million SNPs have been identified in the human genome. Almost all (99%) of these have no biological effect. Over 60,000, however, are within genes, and some are associated with disease. A familiar example is the SNP responsible for the formation of hemoglobin S in sickle cell anemia. SNPs have been classified according to location, relation to coding sequences, and whether they cause a conservative or nonconservative sequence alteration (Table 10.9).

SNP databases such as dbSNP, dbVar, ClinVar, and others are collections of DNA sequence variants used as a reference for screening genomic sequencing data. A variant detected by sequencing may already be described or associated with a disease phenotype as noted in these databases. In addition to SNPs, these databases include short deletions, insertions, and duplications that involve more than one nucleotide.

TABLE 10.9 Types of SNPs

SNP	Region	Alteration
Type I	Coding	Nonconservative
Type II	Coding	Conservative
Type III	Coding	Silent
Type IV	Noncoding 5′ UTR*	
Type V	Noncoding 3′ UTR	
Type VI	Noncoding, other	

*Untranslated region.

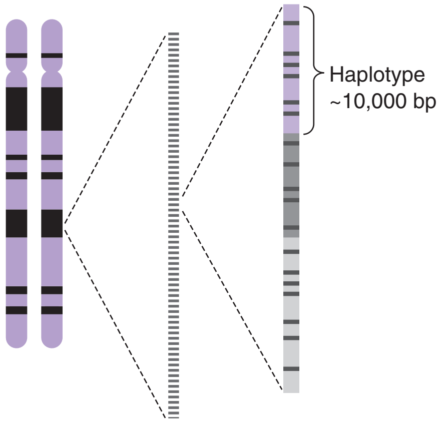


FIGURE 10.25 Sections of DNA along chromosomes can be inherited as a unit or block of sequence in which no recombination occurs. All the SNPs on that block comprise a haplotype.

The Human Haplotype Mapping (HapMap) Project

The international HapMap Project was an initiative started as the Human Genome Project neared completion. Its goal was to develop a haplotype map of the human genome (identify blocks of DNA polymorphisms that are inherited together; Fig. 10.25). This map would then be used to identify common disease associations and patterns of human DNA sequence variation. Although millions of SNPs, RFLPs, VNTRs, and STRs were discovered through the HapMap project, advances in genomic sequencing technology and development of comprehensive population-based databases, such as the 1000 Genomes Project, has supplanted its contributions to research. In 2016, the National Center for Biotechnology Information retired the HapMap resource. Researchers were directed to the 1000 Genomes Project for population genetic and genomic information.

Historical Highlights

The Human Haplotype Mapping project was initiated in October 2002, with a target completion date of September 2005.[48] An initial draft of the HapMap was completed before the deadline date, and a second phase was started to generate an even more detailed map. The new phase increased the density of SNP identification fivefold from 1 SNP per 3,000 bases to 1 SNP per 1,000 bases, or a total of 3.1 million SNPs, approximately 25% of the estimated 11 million SNPs in the human genome.[49] The improved detail of the second-phase project (phase II HapMap) advanced efforts to locate specific genes involved in complex genetic disorders and provided insight into ancestry, recombination, and linkage disequilibrium studies.

MITOCHONDRIAL DNA POLYMORPHISMS

Mitochondria contain a circular genome of 16,569 bp. The two strands of the circular mitochondrial DNA (mtDNA) chromosome have an asymmetric distribution of Gs and Cs generating a G-rich heavy (H) and a C-rich light (L) strand. Each strand is transcribed from a control region starting at one predominant promoter, P_L on the L strand and P_H on the H strand, located in sequences of the mitochondrial circle called the displacement (D)-loop (Fig. 10.25). The D-loop forms a triple-stranded region with a short piece of H-strand DNA, the 7S DNA, synthesized from the H strand. Bidirectional transcription starts from P_L on the L-strand and P_{H1} and P_{H2} on the H-strand. RNA synthesis proceeds around the circle in both directions. A bidirectional attenuator sequence limits L-strand synthesis and, in doing so, maintains a high ratio of rRNA to mRNA transcripts from the H-strand (see Fig. 10.26). Mature mitochondrial RNAs, 1 to 17, are generated by cleavage of the polycistronic (multiple-gene) transcript at the location of the tRNA genes.

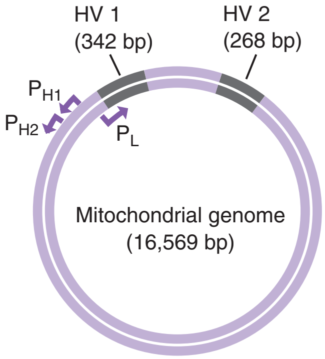


FIGURE 10.26 The mitochondrial genome is circular. The hypervariable (HV) areas in the control region are shown. Mitochondrial genes are transcribed bidirectionally starting at promoters (P_L, P_{H1}, and P_{H2}).

Genes encoded on the mtDNA include 22 tRNA genes, 2 ribosomal RNA genes, and 12 genes coding for components of the oxidation-phosphorylation system. Mutations in these genes are responsible for neuropathies and myopathies. In addition to coding sequences, the mitochondrial genome has two noncoding regions that vary in DNA sequence, the **hypervariable region 1** and the **hypervariable region 2,** or **HV1** and **HV2** (see Fig. 10.27). The reference mtDNA hypervariable region is the sequence published initially by Anderson, called the **Cambridge reference sequence,** the **Oxford sequence,** or the **Anderson reference.**[50] Polymorphisms are denoted as variations from the reference sequence. Nucleotide sequencing of the mtDNA control region has been validated for the genetic characterization of forensic specimens and disease states and for genealogy studies.

In contrast to nuclear DNA, including the Y chromosome, mtDNA follows maternal clonal inheritance patterns. With few exceptions,[51,52] mtDNA types (sequences) are inherited maternally. These characteristics make possible the collection of reference material for forensic analysis, even in cases in which generations are skipped. For forensic purposes, the quality of a match between two mtDNA sources is determined by counting the number of times the mtDNA profile occurs in data collections of unrelated individuals, so the estimate of

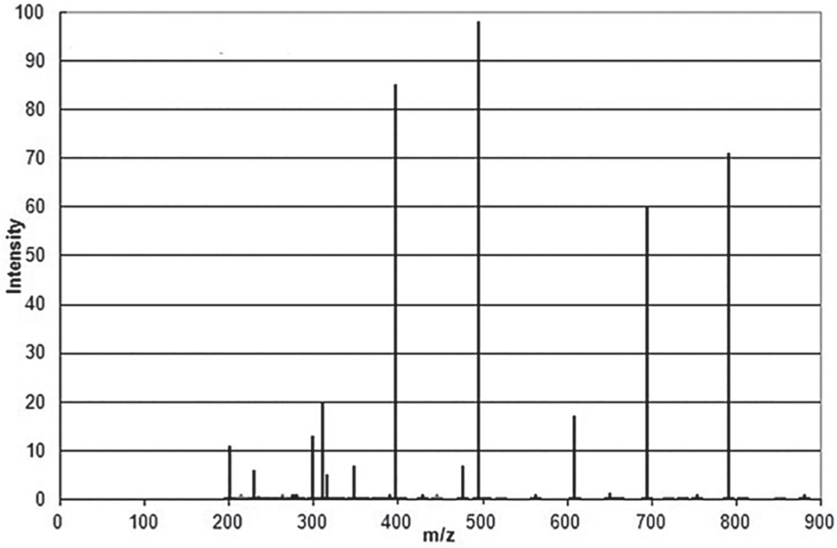


FIGURE 10.27 A peptide spectrum reflects the underlying DNA sequence variants in the form of variant amino acids. Peak patterns (profiles) are compared to reference databases. The probability of a proteomic profile being shared by unrelated individuals is computed from the size of the population contributing spectra to the database. m/z = mass/charge

the uniqueness of a particular mtDNA type depends on the size of the reference database. The more mitochondrial DNA sequences are entered into the database, the more powerful the identification by mitochondrial DNA will become.[53]

Mitochondrial nucleotide sequence data are divided into two components, forensic and public. The forensic component consists of anonymous population profiles and is used to assess the extent of certainty of mtDNA identifications in forensic casework. All forensic profiles include, at a minimum, a sequence region in HV1 (nucleotide positions 16024 to 16383) and a sequence region in HV2 (nucleotide positions 53 to 372). These data are searched through the CODIS program in open case files and missing persons cases. Approximately 610 bp, including the hypervariable regions of mtDNA, are routinely sequenced for forensic analysis. Deviations from the Cambridge reference sequence are recorded as the number of the position and a base designation. For example, a transition from A to G at position 263 would be recorded as 263 G.

The public data consist of mtDNA sequence data from the scientific literature and the GenBank and European Molecular Biology Laboratory databases. The public data have not been subjected to the same quality standards as the forensic data.[54] The public databases provide information on worldwide population groups not contained within the forensic data and can be used for investigative purposes.

As all maternal relatives share mitochondrial sequences, the mtDNA of sisters and brothers or mothers and daughters will exactly match in the hypervariable region in the absence of mutations. Therefore, the use of mtDNA polymorphisms is for exclusion. There is an average of 8.5 nucleotide differences between mtDNA sequences of unrelated individuals in the hypervariable region.

Advanced Concepts

In contrast to nuclear DNA, the human mtDNA genome is completely sequenced and numbered. Variants in the mitochondrial DNA are indicated in relation to the full mitochondrial DNA sequence. Descriptions are preceded by “m.” and reported with the terms used for nuclear DNA (e.g., a T to C change at position 8993 would be m.8993T>C). Descriptions of changes at the protein level include a reference to the protein changed; for example, replacement of leucine with proline at position 156 in ATP synthase 6 would be ATP6:p.Leu156Pro.

The Scientific Working Group for DNA Methods (SWGDAM) and the International Society for Forensic Genetics (ISFG) have recommended guidelines for the use of mtDNA from bone, teeth, or hair for identification purposes.[55,56] The process begins with a visual inspection of the specimen. Bone or teeth specimens are examined and ascertained to be of human origin. In the case of hair samples, the hairs are examined microscopically and compared with hairs from a known source. Sequencing is performed only if the specimen meets the criteria of origin and visual matching to the reference source. Before DNA isolation, the specimens are cleaned with detergent or, for bone or teeth, by sanding to remove any possible source of extraneous DNA adhering to the specimen. The cleaned specimen is then ground in an extraction solution. Hair shafts yield mtDNA, as does the fleshy pulp of teeth or bone. The dentin layer of old tooth samples will also yield mtDNA. DNA is isolated by organic extraction and amplified by PCR. The PCR products are then purified and subjected to dideoxy sequencing. A positive control of a known mitochondrial sequence is included with every run along with a reagent blank for PCR contamination and a negative control for contamination during the sequencing reaction. If the negative or reagent blank controls yield sequences similar to the specimen sequence, the results are rejected. Both strands of the specimen PCR product must be sequenced.

Raw mitochondrial sequence data are imported into a software program for analysis. With the sequence software, the heavy-strand sequences should be reverse-complemented so that the bases are aligned in the light-strand orientation for strand comparison and base designation.

Occasionally, more than one mtDNA population is present in the same individual. This is called **heteroplasmy.** In point heteroplasmy, two DNA bases are observed at the same nucleotide position. Length heteroplasmy is typically a variation in the number of bases in tracts of like bases (homopolymeric tracts, e.g., CCCCC). A length variant alone cannot be used to support an interpretation of exclusion.[57]

In general, if two or more nucleotide differences occur between a reference and a test sample, the test sample can be excluded as originating from the reference or a maternally related person. One nucleotide difference between the samples is interpreted as an inconclusive result. If the test and reference samples show sequence concordance, then the test specimen cannot be excluded as coming from the same individual or maternal relative as the source of the reference sequence. The conclusion that an individual can or cannot be eliminated as a possible source of mtDNA is reached under conditions defined by each individual laboratory. In addition, evaluation of cases in which heteroplasmy may have occurred is laboratory-defined.

The mtDNA profile of a test sample can also be searched in a population database. Population databases, such as the mtDNA population database and CODIS, are used to assess the weight of forensic evidence, based on the number of different mitochondrial sequences previously identified. The quality of sequence information used and submitted for this purpose is extremely important.[58,59] Based on the number of known mtDNA sequences, the probability of sequence concordance in two unrelated individuals is estimated at 0.003. The probability that two unrelated individuals will differ by a single base is 0.014.

Mitochondrial DNA analysis is also used for lineage studies and to track population migrations. As with the Y chromosome, there is no recombination between mitochondria, and polymorphisms arise mostly through mutation. The location and divergence of specific sequences in the HV regions of mitochondria are a historical record of the relatedness of populations.

Because mitochondria are naturally amplified (hundreds per cell and tens of circular genomes per mitochondria) and because of the nuclease- and damage-resistant circular nature of the mitochondrial DNA, mtDNA typing has been a useful complement to other types of DNA identification. Challenging specimens of insufficient quantity or quality for nuclear DNA analysis may still yield useful information from mtDNA. To this end, mtDNA analysis has been helpful for the identification of missing persons in mass disasters or for typing ancient specimens.[60] MtDNA typing can also be applied to quality assurance issues, as described for STR typing of pathology specimens.[61]

OTHER IDENTIFICATION METHODS

Protein-Based Identification

Like DNA, protein contains polymorphisms. Protein polymorphisms are in the form of amino acid sequence variations. Some proteins are chemically more stable than DNA in harsh environments. Proteins, such as keratin and collagen, are also more abundant. It has been proposed that protein polymorphisms may serve as supportive confirmation or even an alternative for DNA identification results.[62]

Nonsynonymous DNA polymorphisms produce single-amino-acid polymorphisms in proteins. A hair shaft, for example, contains over 300 nuclear and mitochondrial proteins, adequately representing the whole genome. A collection of peptide variants comprises a profile. Like STR profiles, peptide profiles could be collected for identification and population-based studies. By associating peptide changes with the known SNP that codes for them, a putative DNA profile might be generated from peptide data where sufficient DNA was not available.

Peptide variants in these proteins can be identified using liquid chromatography followed by mass spectrometry. Proteins isolated from test samples are reduced, alkylated, and digested with trypsin, and the resulting peptides are resolved by liquid chromatography. On-column concentration can increase target molecule concentrations. Particles are then deposited on the matrix, ionized, and subjected to separation by size and charge to generate a spectrum (Fig. 10.27) that can be compared with peptide reference spectra. Software matching algorithms identify peptide variants from the reference spectra.[63] A set of variants is the proteomic profile. This profile can then be compared to a peptide database.[64] There are close to 4 million spectra in available libraries, accounting for different ionization methods. Based on the population frequency of variants (or their associated SNP alleles), the probability of a proteomic profile being shared by unrelated individuals is computed from the size of the population contributing spectra.

Annotated reference peptide libraries from various organisms and proteins are being developed for the rapid matching and identification of acquired peptide spectra from non-human species as well. Among these, smaller, more focused libraries have been collected specifically for humans and mice. The National Institute of Standards and Technology is also developing a peptide mass spectral library to provide peptide reference data for disease-related spectra.

Epigenetic Profiles

DNA and probably protein identification systems cannot distinguish between syngeneic individuals (identical twins). Epigenetic changes occur as a result of environmental events, such that a putative epigenetic profile is unique to each individual because no two individuals will have the same environmental exposures.

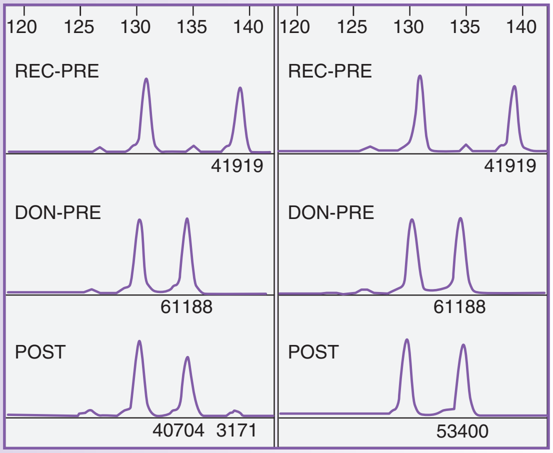
Epigenetic alterations, particularly DNA methylation, change in the absence of cell division or DNA sequence alterations. Many of these changes are stable and can be detected at the DNA sequence level. There are a variety of methods to detect methylated DNA, including methylation-sensitive restriction enzymes, methylation-specific PCR, and bisulfate sequencing by Sanger or massive parallel sequencing.

Although shared epigenetics in families is evidence for the inheritance of epigenetic traits, epigenetic differences due to environmental exposures add an additional level of distinction among individuals. Epigenetic differences, not present at early ages, have been observed in adult identical twins.[65,66] Estimation of the age of persons whose biological materials are recovered at a crime scene is valuable in forensic applications. Strategies have been designed to use epigenetic patterns that accumulate over time to predict chronological age.[67] Through epigenetics, materials commonly found at crime scenes, such as bloodstains, may provide useful information on human age.[68]

Epigenetic markers can also be used to identify body fluids (e.g., saliva, semen, blood). Unique methylation patterns occur at specific gene promoters in these cell types.[69] A marker set defined for fluid type discrimination may also be used to eliminate fluids from other nonprimate species.[70]

Case Study 10.1

A 32-year-old woman was treated for mantle cell lymphoma with a non-myeloablative bone marrow transplant. Before the transplant and after a donor was selected, STR analysis was performed on the donor and the recipient to find informative alleles. One hundred days after the transplant, engraftment was evaluated using the informative STR alleles. The results from one marker, D5S818, are shown in the top panel. One year later, the patient was reevaluated. The results from the same marker are shown in the bottom panel.

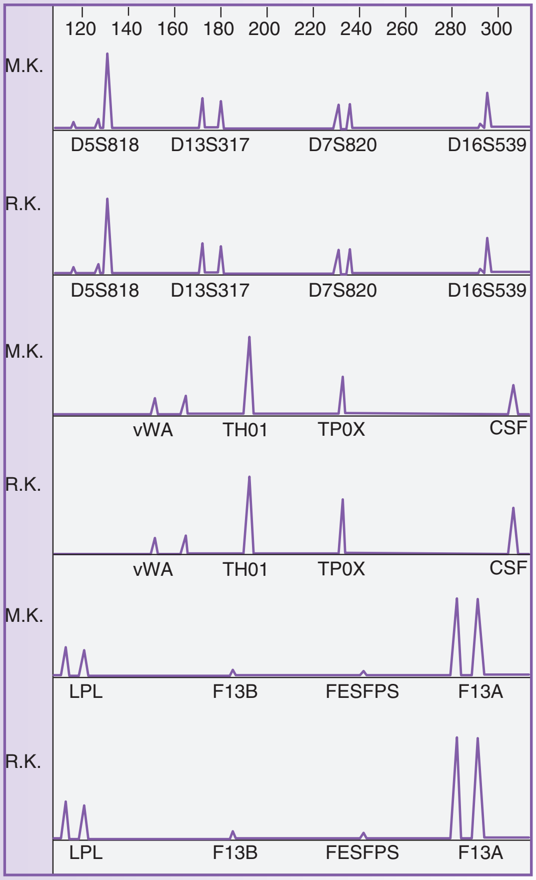


Results from engraftment analysis at 100 days *(top)* and 1 year *(bottom)* showing marker D5S818. REC, recipient; DON, donor.

QUESTION: *Was the woman successfully engrafted with donor cells? Explain your answer.*

Case Study 10.2

A 26-year-old young man reported to his doctor with joint pain and fatigue. Complete blood count and differential counts were indicative of chronic myelogenous leukemia. The diagnosis was confirmed by karyotyping, showing 9/20 metaphases with the t(9;22) translocation. Quantitative PCR was performed to establish a baseline for monitoring tumor load during and following treatment. Treatment with a tyrosine kinase inhibitor and a bone marrow transplant were recommended. The man had a twin brother, who volunteered to donate bone marrow. The two brothers were not sure if they were fraternal or identical twins. Donor and recipient buccal cells were sent to the molecular pathology laboratory for STR informative analysis. The results are shown here.

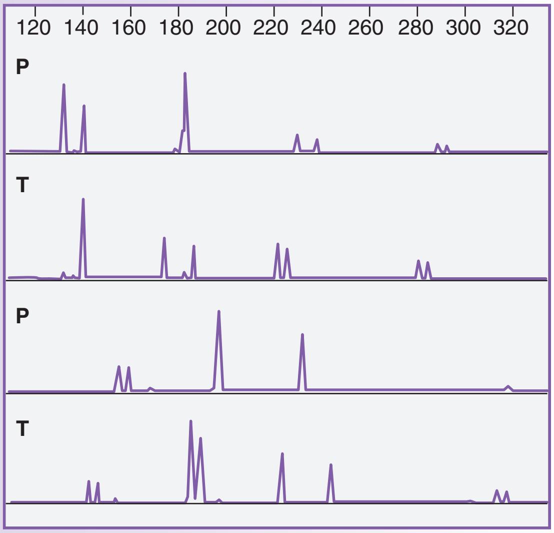


STR analysis of two brothers, one who serves as bone marrow donor (D) to the other (R). Twelve loci are shown.

QUESTION: *Were the brothers fraternal or identical twins? Explain your answer.*

Case Study 10.3

The pathology department received a fixed paraffin-embedded tissue section with a diagnosis of benign uterine fibroids. Slides were prepared for microscopic study. Only benign fibroid cells were observed on all slides, except one. A small malignant process was observed located between the fibroid and normal areas on one slide. Because similar tissue was not observed on any other section, it was possible that the process was a contamination from the embedding process. To determine the origin of the malignant cells, DNA was extracted from the malignant area and compared with DNA extracted from normal tissue from the patient. The results are shown here.



STR analysis of suspicious tissue discovered on a paraffin section. Eight loci were tested. *P,* patient; *T,* tissue section.

QUESTION: *Were the malignant cells seen in one section derived from the patient, or were they a contaminant of the embedding process? Explain your answer.*

STUDY QUESTIONS

1. Consider the following STR analysis.

Locus	Child	Mother	AF1	AF2
D3S1358	15/15	15	15	15/16
vWA	17/18	17	17/18	18
FGA	23/24	22/23	20	24
TH01	6/10	6/7	6/7	9/10
TPOX	11/11	9/11	9/11	10/11
CSF1PO	12/12	11/12	11/13	11/12
D5S818	10/12	10	11/12	12
D13S317	9/10	10/11	10/11	9/11

a. Circle the child's alleles that are inherited from the father.

b. Which alleged father (AF) is not excluded as the biological parent?

2. The following evidence was collected for a criminal investigation.

Locus	Victim	Evidence	Suspect
TPOX	11/12	12, 11/12	11
CSF1PO	10	10, 9	9/10
D13S317	8/10	10, 8/10	9/12
D5S818	9/11	10/11, 9/11	11
TH01	6/10	6/10, 8/10	5/11
FGA	20	20, 20/22	20
vWA	15/17	18, 15/17	15/18
D3S1358	14	15/17, 14	11/12

The suspect is heterozygous at the amelogenin locus.

a. Is the suspect male or female?
b. In the evidence column, circle the alleles belonging to the victim.
c. Should the suspect be held or released?

3. A child and an alleged father (AF) share alleles with the following paternity index.

Locus	Child	AF	Paternity Index for Shared Allele
D5S818	9,10	9	0.853
D8S1179	11	11,12	2.718
D16S539	13,14	10,14	1.782

a. What is the combined paternity index from these three loci?
b. With 50% prior odds, what is the probability of paternity from these three loci?

4. Consider the following theoretical allele frequencies for the loci indicated.

Locus	Alleles	Allele Frequency
CSF1PO	14	0.332
D13S317	9,10	0.210, 0.595
TPOX	8,11	0.489, 0.237

a. What is the overall allele frequency for this genotype, using the product rule?
b. What is the probability that this DNA found at two sources came from the same person?

5. STR at several loci were screened by capillary electrophoresis and fluorescent detection for informative peaks prior to a bone marrow transplant. The following results were observed.

Locus	Donor Alleles	Recipient Alleles
LPL	7,10	7,9
F13B	8,14	8
FESFPS	10	7
F13A01	5,11	5,11

Which loci are informative?

6. An engraftment analysis was performed by capillary gel electrophoresis and fluorescence detection. The fluorescence as measured by the instrument under the FESFPS donor peak was 28,118 units, and that under the FESFPS recipient peak was 72,691. What is the percent donor in this specimen?

7. The T-cell fraction from the blood sample in question 6 was separated and measured for donor cells. Analysis of the FESFPS locus in the T-cell fraction yielded 15,362 fluorescence units under the donor peak and 97,885 under the recipient peak. What does this result predict with regard to T-cell-mediated events such as graft-versus-host disease or graft-versus-tumor?

8. If a child had a Y haplotype including DYS393 allele 12, DYS439 allele 11, DYS445 allele 8, and DYS447 allele 22, what are the predicted Y alleles for these loci of the natural father?

9. Which of these would be used for a surname test: Y-STR, MINI-STR, mitochondrial typing, or autosomal STR?

10. An ancient bone fragment was found and said to belong to an ancestor of a famous family. Living members of the family donated DNA for confirmation of the relationship. What type of analysis would likely be used for this test? Why?

11. What is a biological exception to positive identification by autosomal STR?

12. A partial STR profile was produced from a highly degraded sample. Alleles matched to a reference

sample at five loci. Is this sufficient for a legal identification?

13. What is an SNP haplotype? What are tag SNPs?

14. Which of the following is an example of linkage disequilibrium?
    a. Seventeen members of a population of 1,000 people have a rare disease, and all 17 people have the same haplotype at a particular genetic location on chromosome 3.
    b. Five hundred people from a population of 1,000 people have the same SNP on chromosome 3.

15. Why are SNPs superior to STR and RFLP for mapping and association studies?

References

1. Lander E, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature* 2001;409:860–921.
2. Herrin GJ. Probability of matching RFLP patterns from unrelated individuals. *American Journal of Human Genetics* 1993;52:491–497.
3. Hall J, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, King MC. Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* 1990;250:1684–1689.
4. King M-C. Localization of the early-onset breast cancer gene. *Hospital Practice*, 1991;26:89–94.
5. Gill P, Jeffreys AJ, Werrett DJ. Forensic applications of DNA "fingerprints." *Nature* 1985;318:577–579.
6. Budowle B, Bacchtel FS. Modifications to improve the effectiveness of restriction fragment length polymorphism typing. *Applied and Theoretical Electrophoresis* 1990;1:181–187.
7. Evett I, Gill P. A discussion of the robustness of methods for assessing the evidential value of DNA single locus profiles in crime investigations. *Electrophoresis* 1991;12:226–230.
8. Walsh P, Fildes N, Louie AS, Higuchi R. Report of the blind trial of the Cetus Amplitype HLA DQ alpha forensic deoxyribonucleic acid (DNA) amplification and typing kit. *Journal of Forensic Science* 1991;36:1551–1556.
9. Butler J. *Forensic DNA typing: biology and technology behind STR markers*. London: Academic Press; 2001.
10. Much M, Buza, N, Hui P. Tissue identity testing of cancer by short tandem repeat polymorphism: pitfalls of interpretation in the presence of microsatellite instability. *Human Pathology* 2014;45:549–555.
11. Turrina S, Atzei R, Filippini G, De Leo D. STR typing of archival Bouin's fluid-fixed paraffin-embedded tissue using new sensitive redesigned primers for three STR loci (CSF1P0, D8S1179 and D13S317). *Journal of Forensic and Legal Medicine* 2008;15:27–31.
12. Marjanović D, Durmić-Pasić A, Kovacević L, Avdić J, Dzehverović M, Haverić S, Ramić J, Kalamujić B, Lukić Bilela L, Skaro V, Projić P, Bajrović K, Drobnic K, Davoren J, Primorac D. Identification of skeletal remains of Communist Armed Forces victims during and after World War II: Combined Y-chromosome (STR) and MiniSTR approach. *Croation Medical Journal* 2009;50:296–304.
13. Kayser M, Caglià A, Corach D, Fretwell N, Gehrig C, Graziosi G, Heidorn F, Herrmann S, Herzog B, Hidding M, Honda K, Jobling M, Krawczak M, Leim K, Meuser S, Meyer E, Oesterreich W, Pandya A, Parson W, Penacino G, Perez-Lezaun A, Piccinini A, Prinz M, Schmitt C, Roewer L, et al. Evaluation of Y-chromosomal STRs: a multicenter study. *International Journal of Legal Medicine* 1997;110:141–149.
14. Leclair B, Frageau CJ, Bowen KL, Fourney RM. Precision and accuracy in fluorescent short tandem repeat DNA typing: assessment of benefits imparted by the use of allelic ladders with the AmpF/STR Profiler Plus kit. *Electrophoresis* 2004;25:790–796.
15. Olaisen B, Bär W, Brinkmann B, Budowle B, Carracedo A, Gill P, Lincoln P, Mayr WR, Rand S. DNA Recommendations 1997 of the International Society for Forensic Genetics. *Vox Sanguinis* 1998;74:61–63.
16. Butler E, Li R. Genetic markers for sex identification in forensic DNA analysis. *Journal of Forensic Investigation* 2014;2:1–10.
17. Morin P, Manaster C, Mesnick SL, Holland R. Normalization and binning of historical and multi-source microsatellite data: overcoming the problems of allele size shift with allelogram. *Molecular Ecology Resources* 2009;9:1451–1455.
18. Monson K, Budowle B. A comparison of the fixed bin method with the floating bin and direct count methods: effect of VNTR profile frequency estimation and reference population. *Journal of Forensic Science* 1993;38:1037–1050.
19. Norton H, Neel JV. Hardy-Weinberg equilibrium and primitive populations. *American Journal of Human Genetics* 1965;17:91–92.
20. Hartmann J, Houlihan BT, Keister RS, Buse EL. The effect of ethnic and racial population substructuring on the estimation of multi-locus fixed-bin VNTR RFLP genotype probabilities. *Journal of Forensic Science* 1997;42:232–240.
21. Tvedebrink T, Bright JA, Buckleton JS, Curran JM, Morling N. The effect of wild card designations and rare alleles in forensic DNA database searches. *Forensic Science International Genetics* 2015;16:98–104.
22. Gill P, Whitaker J, Flaxman C, Brown N, Buckleton J. An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA. *Forensic Science International* 2000;112:17–40.
23. Gilbert N. DNA's identity crisis. *Nature* 2010;464:347–348.
24. Wenk R, Traver M, Chiafari F. Determination of sibship in any two persons. *Transfusion* 1996;36:259–262.
25. Douglas J, Boehnke M, Lange K. A multipoint method for detecting genotyping errors and mutations in sibling-pair linkage data. *American Journal of Human Genetics* 2000;66:1287–1297.
26. Huang J, Huang Y. Sibship analysis based on parental genotype reconstruction from any number of reference siblings. *Journal of the Forensic Science Society* 2017;62:233–238.

27. Epstein M, Duren WL, Boehnke M. Improved inference of relationship for pairs of individuals. *American Journal of Human Genetics* 2000;67:1219–1231.

28. Kayser M. Forensic use of Y-chromosome DNA: a general overview. *Human Genetics* 2017;136:621–635.

29. Helgason, A, Einarsson AW, Guðmundsdóttir VD, Sigurðsson A, Gunnarsdóttir ED, Jagadeesan A, Ebenesersdóttir S, Kong A, Stefansson K. The Y-chromosome point mutation rate in humans. *Nature Genetics* 2015;47:453–457.

30. Roewer L, Krawczak M, Willuweit S, Nagy M, Alves C, Amorim A, Anslinger K, Augustin C, Betz A, Bosch E, Caglia A, Carracedo A, Corach D, Dekairelle AF, Dobosz T, Dupuy BM, Furedi S, Gehrig C, GusmaÃμ L, Henke J, Henke L, Hidding M, Hohoff C, Hoste B, Jobling MA, Kargel HJ, de Knijff P, Lessig R, Liebeherr E, Lorente M, Martinez-Jarreta B, Nievas P, Nowak M, Parson W, Pascali VL, Penacino G, Ploski R, Rolf B, Sala A, Schmidt U, Schmitt C, Schneider PM, Szibor R, Teifel-Greding J, Kayser M. Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes. *Forensic Science International* 2001;118:106–113.

31. Van Deerlin V, Leonard DGB. Bone marrow engraftment analysis after allogeneic bone marrow transplantation. *Acute Leukemias* 2000;20:197–225.

32. Thiede C, Florek M, Bornhauser M, Ritter M, Mohr B, Brendel C, Ehninger G, Neubauer A. Rapid quantification of mixed chimerism using multiplex amplification of short tandem repeat markers and fluorescence detection. *Bone Marrow Transplant* 1999;23:1055–1060.

33. Smith A, Martin PJ. Analysis of amplified variable number tandem repeat loci for evaluation of engraftment after hematopoietic stem cell transplantation. *Reviews in Immunogenetics* 1999;1:255–264.

34. Oliver D, Thompson RE, Griffin CA, Eshleman JR. Use of single nucleotide polymorphisms (SNP) and real-time polymerase chain reaction for bone marrow engraftment analysis. *Journal of Molecular Diagnostics* 2000;2:202–208.

35. Thiede C, Bornhauser M, Ehninger G. Evaluation of STR informativity for chimerism testing - comparative analysis of 27 STR systems in 203 matched related donor recipient pairs. *Leukemia* 2004;18:248–254.

36. Clark J, Scott SD, Jack AL, Lee H, Mason J, Carter GI, Pearce L, Jackson T, Clouston H, Sproul A, Keen L, Molloy K, Folarin N, Whitby L, Snowden JA, Reilly JT, Barnett D. United Kingdom National External Quality Assessment Service for Leucocyte Immunophenotyping Chimerism Working Group: monitoring of chimerism following allogeneic haematopoietic stem cell transplantation (HSCT): technical recommendations for the use of short tandem repeat (STR) based techniques, on behalf of the United Kingdom National External Quality Assessment Service for Leucocyte Immunophenotyping Chimerism Working Group. *British Journal of Haematology* 2015;168:26–37.

37. Platt E, Sommer P, McDonald L, Bennett A, Hunt J. Tissue floaters and contaminants in the histology laboratory. *Archives of Pathology & Laboratory Medicine* 2009;133:973–978.

38. Hunt J, Swalsky P, Sasatomi E, Niehouse L, Bakker A, Finkelstein SD. A microdissection and molecular genotyping assay to confirm the identity of tissue floaters in paraffin-embedded tissue blocks. *Archives of Pathology & Laboratory Medicine* 2003;127:213–217.

39. Junge A, Dettmeyer R, Madea B. Identification of biological samples in a case of contamination of a cytological slide preparation. *Journal of Forensic Science* 2008;53:739–741.

40. Bossuyt V, Buza N, Ngo NT, Much MA, Asis MC, Schwartz PE, Hui P. Cancerous "floater": a lesson learned about tissue identity testing, endometrial cancer and microsatellite instability. *Modern Pathology* 2013;26:1264–1269.

41. Much M, Buza N, Hui P. Tissue identity testing of cancer by short tandem repeat polymorphism: pitfalls of interpretation in the presence of microsatellite instability. *Human Pathology* 2014;45: 549–555.

42. Kruglyak L, Nickerson DA. Variation is the spice of life. *Nature Genetics* 2001, 27:234–236.

43. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, Hunt SE, Cole CG, Coggill PC, Rice CM, Ning Z, Rogers J, Bentley DR, Kwok PY, Mardis ER, Yeh RT, Schultz B, Cook L, Davenport R, Dante M, Fulton L, Hillier L, Waterston RH, McPherson JD, Gilman B, Schaffner S, Van Etten WJ, Reich D, Higgins J, Daly MJ, Blumenstiel B, Baldwin J, Stange-Thomann N, Zody MC, Linton L, Lander ES, Altshuler D; International SNP Map Working Group. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 2001;409:928–933.

44. Wheeler DA SM, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ, Makhijani V, Roth GT, Gomes X, Tartaro K, Niazi F, Turcotte CL, Irzyk GP, Lupski JR, Chinault C, Song XZ, Liu Y, Yuan Y, Nazareth L, Qin X, Muzny DM, Margulies M, Weinstock GM, Gibbs RA, Rothberg JM. The complete genome of an individual by massively parallel DNA sequencing. *Nature* 2008;452:872–876.

45. Matise T, Sachidanandam R, Clark AG, Kruglyak L, Wijsman E, Kakol J, Buyske S, Chui B, Cohen P, de Toma C, Ehm M, Glanowski S, He C, Heil J, Markianos K, McMullen I, Pericak-Vance MA, Silbergleit A, Stein L, Wagner M, Wilson AF, Winick JD, Winn-Deen ES, Yamashiro CT, Cann HM, Lai E, Holden AL. A 3.9-centimorgan-resolution human single-nucleotide polymorphism linkage map and screening set. *American Journal of Human Genetics* 2003;73:271–284.

46. Cheng Y, Mailund T, Nielsen R. Fast admixture analysis and population tree estimation for SNP and NGS data. *Bioinformatics* 2017;33(14):2148–2155.

47. De Wit P. SNP discovery using next generation transcriptomic sequencing. *Methods in Molecular Biology* 2016;1452: 81–95.

48. Consortium TIH. The International HapMap Project. *Nature* 2003;426:789–796.

49. Consortium TIH. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007;449:851–862.

50. Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG. Sequence and organization of the human mitochondrial genome. *Nature* 1981;290: 457–465.

51. Schwartz M, Vissing J. Paternal inheritance of mitochondrial DNA. *New England Journal of Medicine* 2002;347: 576–580.

52. Yu Z, O'Farrell PH, Yakubovich N, DeLuca SZ. The mitochondrial DNA polymerase promotes elimination of paternal mitochondrial genomes. *Current Biology* 2017;27:1033–1039.

53. Lee Y, Kim WY, Ji M, Kim JH, Bhak J. MitoVariome: a variome database of human mitochondrial DNA. *BMC Genomics* 2009; 10:S12.

54. Melton T, Holland C, Holland M. Forensic mitochondrial DNA analysis: current practice and future potential. *Forensic Science Reviews* 2012;24:101–122.

55. (SWGDAM) SWGoDAM. Guidelines for mitochondrial DNA (mtDNA) nucleotide sequence interpretation. *Forensic Science Communications* 2003;5:1–5.

56. Parson W, Gusmão L, Hares DR, Irwin JA, Mayr WR, Morling N, Pokorak E, Prinz M, Salas A, Schneider PM, Parsons TJ. DNA Commission of the International Society for Forensic Genetics: revised and extended guidelines for mitochondrial DNA typing. *Forensic Science International Genetics* 2014;13:134–142.

57. Stewart JE, Fisher CL, Aagaard PJ, Wilson MR, Isenberg AR, Polanskey D, Pokorak E, DiZinno JA, Budowle B. Length variation in HV2 of the human mitochondrial DNA control region. *Journal of Forensic Science* 2001;46:862–870.

58. Bandelt H-J, Salas A, Bravi C. Problems in FBI mtDNA database. *Science* 2004;305:1402–1404.

59. Budowle B, Polanskey D. FBI mtDNA database: a cogent perspective. *Science* 2005;307:845–847.

60. Zupanic I, Gornjak Pogorelc B, Balazic J. Molecular genetic identification of skeletal remains from the Second World War Konfin I mass grave in Slovenia. *International Journal of Legal Medicine* 2010;124:307–317.

61. Alonso A, Alves C, Suárez-Mier MP, Albarrán C, Pereira L, de Simó LF, Martín P, García O, Gusmão L, Sancho M, Amorim A. Mitochondrial DNA haplotyping revealed the presence of mixed up benign and neoplastic tissue sections from two individuals on the same prostatic biopsy slide. *Journal of Clinical Pathology* 2005;58:83–86.

62. Parker G, Leppert T, Anex DS, Hilmer JK, Matsunami N, Baird L, Stevens J, Parsawar K, Durbin-Johnson BP, Rocke DM, Nelson C, Fairbanks DJ, Wilson AS, Rice RH, Woodward SR, Bothner B, Hart BR, Leppert M. Demonstration of protein-based human identification using the hair shaft proteome. *PLoS One* 2016;11:e0160653.

63. Horlacher O, Nikitin F, Alocci D, Mariethoz J, Müller M, Lisacek F. MzJava: an open source library for mass spectrometry data processing. *Journal of Proteomics* 2015;129:63–70.

64. Yilmaz Ş, Vandermarliere E, Martens L. Methods to calculate spectrum similarity. *Methods in Molecular Biology* 2017;1549:75–100.

65. Fraga M, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, Heine-Suñer D, Cigudosa JC, Urioste M, Benitez J, Boix-Chornet M, Sanchez-Aguilera A, Ling C, Carlsson E, Poulsen P, Vaag A, Stephan Z, Spector TD, Wu YZ, Plass C, Esteller M. Epigenetic differences arise during the lifetime of monozygotic twins. *Proceedings of the National Academy of Sciences* 2005;102:10604–10609.

66. Xu J, Fu G, Yan L, Craig JM, Zhang X, Fu L, Ma C, Li S, Cong B. LINE-1 DNA methylation: a potential forensic marker for discriminating monozygotic twins. *Forensic Science International Genetics* 2015;19:136–145.

67. Vidaki A, Ballard D, Aliferi A, Miller TH, Barron LP. Syndercombe Court D: DNA methylation-based forensic age prediction using artificial neural networks and next generation sequencing. *Forensic Science International Genetics* 2017;28:225–236.

68. Zubakov D, Liu F, Kokmeijer I, Choi Y, van Meurs JB, van IJcken WF, Uitterlinden AG, Hofman A, Broer L, van Duijn CM, Lewin J, Kayser M. Human age estimation from blood using mRNA, DNA methylation, DNA rearrangement, and telomere length. *Forensic Science International Genetics* 2016; 24:33–43.

69. Antunes J, Silva DS, Balamurugan K, Duncan G, Alho CS, McCord B. High-resolution melt analysis of DNA methylation to discriminate semen in biological stains. *Analytical Biochemistry* 2016;494:40–45.

70. Silva D, Antunes J, Balamurugan K, Duncan G, Alho CS, McCord B. Developmental validation studies of epigenetic DNA methylation markers for the detection of blood, semen and saliva samples. *Forensic Science International Genetics* 2016;23:55–63.

71. Bar W, Brinkman B, Budowle B, Carracedo A, Gill P, Lincoln P, Mayr W, Olaisen B. DNA recommendations. Further report of the DNA Commission of the ISFH regarding the use of short tandem repeat systems. International Society for Forensic Haemogenetics. *International Journal of Legal Medicine* 1997;110:175–176.

72. Sinha S, Budowle B, Chakraborty R, Paunovic A, Guidry RD, Larsen C, Lal A, Shaffer M, Pineda G, Sinha SK, Schneida E, Nasir H, Shewale JG. Utility of the Y-STR typing systems Y-PLEXTM 6 and Y-PLEXTM 5 in forensic casework and 11 Y-STR haplotype database for three major population groups in the United States. *Journal of Forensic Science* 2004;49:1–10.

73. Iida R, Sawazaki K, Ikeda H, Miyamoto T, Tsubota E, Takatsuka H, Masuyama M, Matsuki T, Yasuda T, Kishi K. A novel multiplex PCR system consisting of Y-STRs DYS441, DYS442, DYS443, DYS444, and DYS445. *Journal of Forensic Science* 2003;48:1088–1090.

74. Hanson E, Ballantyne J. A highly discriminating 21 locus Y-STR "megaplex" system designed to augment the minimal haplotype loci for forensic case work. *Journal of Forensic Science* 2004;49:1–12.

75. Schoskea R, Vallonea PM, Klinea MC, Redmana JW, Butler JM. High-throughput Y-STR typing of U.S. populations with 27 regions of the Y chromosome using two multiplex PCR assays. *Forensic Science International* 2004;139:107–121.

Chapter 11

Detection and Identification of Microorganisms

Outline

- SPECIMEN COLLECTION
- SAMPLE PREPARATION
- QUALITY ASSURANCE
    - *Controls*
    - *Quality Control*
    - *Selection of Sequence Targets for Detection of Microorganisms*
- MOLECULAR DETECTION OF MICROORGANISMS
    - *Bacteria*
        - Respiratory Tract Pathogens
        - Urogenital Tract Pathogens
    - *Viruses*
        - Mass Spectrometry
    - *Mycology*
    - *Parasites*
- ANTIMICROBIAL AGENTS
    - *Resistance to Antimicrobial Agents*
    - *Molecular Detection of Resistance*
        - Beta-Lactam Antibiotic Resistance
        - Glycopeptide Antibiotic Resistance
        - Antimicrobial Resistance in *M. tuberculosis*
- MOLECULAR EPIDEMIOLOGY
    - *Molecular Strain Typing Methods for Epidemiological Studies*
        - Plasmid Analysis
        - Pulsed-Field Gel Electrophoresis
        - Restriction Fragment Length Polymorphism Analysis
        - Arbitrarily Primed PCR
        - Amplified Fragment Length Polymorphism (AFLP) Assay
        - Interspersed Repetitive Elements
        - Internal Transcribed Spacer Elements
        - spa Typing
        - Multilocus Sequence Typing
        - Mass Spectrometry
    - *Comparison of Typing Methods*

Objectives

11.1 Name the organisms that are common targets for molecular-based laboratory tests.

11.2 Identify advantages and disadvantages of using molecular-based methods as compared with traditional culture-based methods in the detection and identification of microorganisms.

11.3 Differentiate between organisms for which commercially available nucleic acid amplification tests exist and those for which **"home-brew"** polymerase chain reaction (PCR) is used.

11.4 List the genes involved in the emergence of antimicrobial resistance that can be detected by nucleic acid amplification methods.

11.5 Compare and contrast the molecular methods that are used to type bacterial strains in epidemiological investigations.

11.6 Explain the value of controls, in particular, amplification controls, in ensuring the reliability of PCR results.

11.7 Interpret pulsed-field gel electrophoresis patterns to determine whether two isolates are related to or different from each other.

Microbiological applications for the clinical laboratory are increasingly based on the molecular characterization of microorganisms and the development and evaluation of molecular-based laboratory tests of clinical specimens isolated in cultures. Another important application of molecular technology in the clinical microbiology laboratory is in the comparison of biochemically similar organisms in outbreak situations, known as molecular epidemiology, to ascertain whether the isolates have a common or independent source. Clinically important microorganisms include a range of life-forms, from arthropods to prions, and although molecular-based methods have become routine in clinical microbiology, traditional culture and biochemical testing are still important for the detection and identification of a variety of microorganisms.

In contrast to the analysis of phenotypic traits (microscopic and colonial morphologies, enzyme or pigment production, carbohydrate fermentation patterns), the analyte for molecular testing is the genome, transcriptome, or proteome of the microorganism. Bacteria, fungi, and parasites have DNA genomes, whereas viruses can have DNA or RNA genomes. Prions, which cause transmissible encephalopathies such as Creutzfeldt–Jakob disease, consist only of protein.

Microorganisms targeted by molecular-based laboratory tests have been those that are difficult and/or time-consuming to isolate, such as *Mycobacterium tuberculosis* as well as other species of *Mycobacterium*; those that are hazardous with which to work in the clinical laboratory, such as *Histoplasma* and *Coccidioides*; and those for which reliable laboratory tests were lacking, such as hepatitis C virus (HCV) and human immunodeficiency virus (HIV).

Additionally, molecular-based tests have been developed for organisms that are received in clinical laboratories in high volumes, such as *Streptococcus pyogenes* in throat swabs and *Neisseria gonorrhoeae* and *Chlamydia trachomatis* in genital specimens. Furthermore, genes that confer resistance to antimicrobial agents are the targets of molecular-based methodologies, such as *mecA*, which contributes to the resistance of *Staphylococcus aureus* to oxacillin; *vanA*, *vanB*, and *vanC,* which give *Enterococcus* resistance to vancomycin; *tonB,* which confers resistance to carbapenems; and *katG* and *inhA,* which mediate *M. tuberculosis* resistance to isoniazid. Furthermore, mass spectrometry can directly identify resistance factors such as expressed β-lactamases even in the absence of antibiotics.[1]

Finally, characterization of DNA, RNA, and protein was developed to find and identify new organisms and to further characterize or classify known organisms, such as influenza virus.[2] Nucleic acid sequence information is used to reclassify bacterial organisms based on 16S rRNA sequence homology, for epidemiological purposes, and to predict therapeutic efficacy. Mass spectrometry is also being applied to the identification of microorganisms based on peptide profiles.

The molecular methods used in the clinical microbiology laboratory are the same as those that were described previously for the identification of human polymorphisms and those that will be discussed in subsequent chapters for the identification of genes involved in cancer and in inherited diseases. These include polymerase chain reaction (PCR)—traditional, real-time, and reverse transcriptase PCR, and DNA sequencing. Additional methods used in molecular epidemiology are pulsed-field gel electrophoresis (PFGE), matrix-assisted laser desorption ionization (MALDI) spectrometry, and other methods that are discussed in this chapter. The development of molecular-based methods has been successful for some organisms but not yet for all organisms, as discussed in this chapter.

SPECIMEN COLLECTION

As with any clinical test, proper procedure is important for collection and transport of specimens for infectious disease testing. Microbiological specimens may require special handling to preserve the viability of the target organism. Special collection systems have been designed for the collection of strict anaerobes, viruses, and other fastidious organisms. Although viability is not as critical for most molecular testing, the quality of nucleic acids may be compromised if the specimen is improperly handled. DNA and especially RNA will be damaged in lysed or nonviable cells. Due to the sensitivity of molecular testing, it is also important to avoid contamination that could yield false-positive results.

Collection techniques designed to avoid contamination from the surrounding environment of adjacent tissues apply to molecular testing, especially to those tests that use amplification methods. Sampling must include material from the original infection. The time and site of collection should be optimal for the likely presence of the infectious agent. For example, *Salmonella typhi* is initially present in peripheral blood but not in urine or stool until at least 2 weeks after infection. For classical methods that include culturing of the agent, a sufficient number of microorganisms must be obtained for agar or liquid culture growth. For molecular testing, however, minimum numbers (as few as 50 organisms) can be detected successfully. The quantity of target organisms, as well as the clinical implications, should be taken into account when interpreting the significance of positive results. Molecular detection can reveal infective agents at levels below clinical significance. Conversely, highly specific molecular methods may miss detection of a variant organism.

Equipment and reagents used for specimen collection are also important for molecular testing (Table 11.1). Blood draws should go into the proper anticoagulant, if one is to be used. Although wooden-shafted swabs may be used for throat cultures, Dacron or calcium alginate swabs with plastic shafts have been recommended for collection of bacteria, viruses, and mycoplasma from mucosal surfaces.[3] The plastics are less adherent to the microorganisms and will not interfere with PCR reagents, with the exception of calcium alginate swabs with aluminum shafts, which had been reported to affect PCR amplification.[4] Collection media or amplification system may influence positivity rates.[5] Collection methods such as Swab Extraction Tube System (SETS), sonication, and vortex have been designed for maximum recovery of microorganisms from swabs by centrifugation[6] (Fig. 11.1).

TABLE 11.1 Specimen Transport Systems

Type	Examples
Sterile containers	Sterile cups, screw-capped tubes, stoppered tubes, Petri dishes
Swabs	Calcium alginate swabs, Dacron swabs, cotton swabs, nasopharyngeal-urogenital swabs, swab transport system
Specialty systems	*Neisseria gonorrhoeae* transport systems, Swab Extraction Tube System (SETS)
Proprietary systems	Molecular testing, *N. gonorrhoeae* transport systems, STAR buffer[42]
Anaerobic transport systems	Starplex Anaerobic Transport system (Fisher), BBL Vacutainer Anaerobic Specimen Collector
Viral transport systems	BD Cellmatics Viral Transport Pack, BBL Viral Culturette

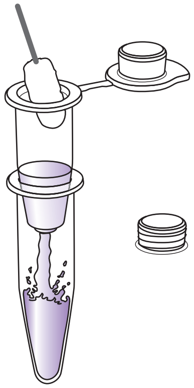


FIGURE 11.1 The Swab Extraction Tube System (SETS) consists of a punctured inner tube that holds a swab and fits in a capable outer tube. Under the force of centrifugation, liquid in the swab is forced through the inner tube and into the outer tube, where it can be stored.

Commercial testing kits supply an optimized collection system for a particular test organism. The Clinical and Laboratory Standards Institute has published documents addressing the requirements for transport devices and quality control guidelines. The College of American Pathologists requires documented procedures describing specimen handling, collection, and transport in each laboratory.

SAMPLE PREPARATION

Isolating nucleic acids from microorganisms is similar to isolating nucleic acids from human cells with only a few additional considerations. First, depending on the microorganism, more rigorous lysis procedures may be required. Mycobacteria and fungi, in particular, have thick cell walls that are more difficult to lyse than those of other bacteria and parasites. Gram-positive bacteria has a thicker cell wall than gram-negative bacteria and may require more rigorous cell lysis conditions. *Mycoplasma,* on the other hand, lacks a cell wall, and so care must be taken with the sample to avoid spontaneous lysis of the cells and loss of nucleic acids.

Advanced Concepts

Biological safety is an important concern for clinical microbiology. Because various collection, transport, and extraction systems inactivate organisms at different times, the technologist should follow the recommendations of the Centers for Disease Control and Prevention (CDC) that call for universal precautions, treating all specimens as if they were infectious throughout the extraction process. Updated guidelines are available from the CDC for the handling of suspected bioterrorism material. Organisms such as smallpox must be handled only in approved (level 4 containment) laboratories. Molecular testing has eased the requirements for laboratory culture. Methods devised to replace the growth of cultures should improve safety levels.

Second, the concentration of organisms within the clinical sample must be considered. Samples can be centrifuged to concentrate the organisms within the fluid from the milliliters of sample that are often received down to volumes that are appropriate for molecular procedures. Third, inhibitors of enzymes used in molecular analysis may be present in clinical specimens; removal or inactivation of inhibitors might be included in specimen preparation methods. Finally, if RNA is to be analyzed, inactivation or removal of RNases in the sample and in all reagents and materials that come into contact with the sample is important.

Any clinical specimen can be used as a source of microorganism nucleic acid for analysis. Depending on the specimen, however, special preparation procedures may be necessary to allow for optimal nucleic acid isolation, amplification, and analysis. The presence of inhibitors of DNA polymerase has been demonstrated in clinical samples; therefore, careful separation of nucleic acid from other molecules present in the sample will ensure target amplification.[7] When processing a whole-blood specimen, it is important to remove hemoglobin and other products of metabolized hemoglobin because they can inhibit DNA polymerase and thus may prevent the amplification of nucleic acid in the sample, resulting in a false-negative PCR result. Eukaryotic cells can be used as a source of nucleic acid for organisms, primarily viruses that infect these cells. In blood samples, white blood cells are isolated from the red blood cells using Ficoll-Hypaque and then lysed. Alternatively, whole blood is processed in automated DNA isolation systems, which effectively remove hemoglobin and other contaminating molecules. Serum and plasma (devoid of red blood cells) are also used as sources of microorganism nucleic acid.

Sputum is a source of nucleic acid from organisms that cause respiratory tract infections. Acidic polysaccharides present in sputum may inhibit DNA polymerase and thus must be removed. Using a method or an instrument that reliably separates DNA from other cellular molecules is sufficient to remove the inhibitors. Urine, when sent for nucleic acid isolation and amplification, is treated similarly to cerebrospinal fluid; that is, the specimen is centrifuged to concentrate the organisms and then subjected to nucleic acid isolation procedures. Inhibitors of DNA polymerase—nitrate, crystals, hemoglobin, and beta-human chorionic gonadotropin—have been demonstrated in urine as well.[8]

The type of specimen used for molecular testing will also affect extraction and yield of nucleic acid. For example, viral nucleic acid from plasma is easier to isolate than nucleic acid from pathogenic *Enterococcus* in stool specimens. Reagents and devices have been developed to combine collection and extraction of nucleic acid from difficult specimens; for example, stool transport and recovery (STAR) buffer or the FTA paper systems that inactivate infectious agents and adhere nucleic acids to magnetic beads or paper, respectively.

QUALITY ASSURANCE

For any type of medical laboratory procedure, quality control is critical for ensuring the accuracy of patient results. Ensuring the quality of the molecular methods is equally important. Molecular testing sensitivity is relatively high, so even one molecule of target is a potential template. Thus, the integrity of specimens, that is, specimens not contaminated by other organisms or with the products of previous amplification procedures, is critical to avoid inaccurate results. Also, it is equally important to ensure that the lack of a product in an amplification procedure is due to the absence of the target organism and not the presence of inhibitors preventing the amplification of target sequences.

Controls

Control substances of known composition are used to monitor the reliability of the method and the input specimen material. The incorporation of positive controls shows that an assay system is functioning properly. A sensitivity control that is positive at the lower **limit of detection** demonstrates the sensitivity of qualitative assays. Two positive controls, one at the lower limit and the other at the upper limit of detection, are run in quantitative assays to test the dynamic range of the assay. Reagent blank or contamination controls are critical for monitoring reagents for contamination; the latter contains all of the assay reagents except target sequences and should always be negative. For typing and other studies that might include nontarget organisms, a negative template control might also be included. This control will detect the presence of the unwanted target(s), but should not react with the desired target.

With regard to amplification methods that are interpreted by the presence or absence of product, false-negative results can occur due to amplification failure. In order to rule out this type of false-negative result, an amplification control aimed at a target that is always present can be incorporated into an amplification assay. The negative template control sample should have a positive amplification control signal, whereas the reagent blank should be negative for amplification. With a positive amplification control, lack of amplification of the target can be more confidently interpreted as a **true negative** result. Amplification controls are usually housekeeping genes or those that are always present in human or microbiological samples. Housekeeping genes that are used as internal controls include prokaryotic genes, such as *groEL, rpoB, recA,* and *gyrB,* and eukaryotic genes, such as β-actin, glyceraldehyde-3-phosphate, interferon-γ, extrinsic homologous control, human mitochondrial DNA, and peptidylprolyl isomerase A.[9]

Internal controls are amplification controls that monitor particular steps of an amplification method. They can be either **homologous extrinsic, heterologous extrinsic,** or **heterologous intrinsic** (Fig. 11.2). A homologous extrinsic control is a target-derived control with a non-target-derived sequence insert. This control is added to every sample after nucleic acid extraction and before amplification. The amplification of this control occurs using the same primers as for the target, which is good for ensuring that amplification occurs in the sample but does not control for target nucleic acid degradation during extraction. Heterologous extrinsic controls are non-target-derived controls that are added to every sample before nucleic acid extraction. This control will ensure that the extraction and amplification procedures were acceptable, but a second set of primers must also be added to the reaction for the control to be amplified. Using this control requires that the procedure be optimized so that the amplification of the control does not interfere with the amplification of the target. Heterologous intrinsic controls are nontarget sequences naturally present in the sample, such as eukaryotic genes in a test for microorganisms. In this case, human gene controls serve to ensure that human nucleic acid is present in the sample in addition to controlling for extraction and amplification. Using this control requires that either two amplification reactions be performed on the sample, one for the control and the other for the target gene, or that the amplification procedure be multiplexed, as long as there is no interference with the amplification of the target.

Quality Control

In a procedure that detects a microorganism, a positive result states that the organism is present in that sample,

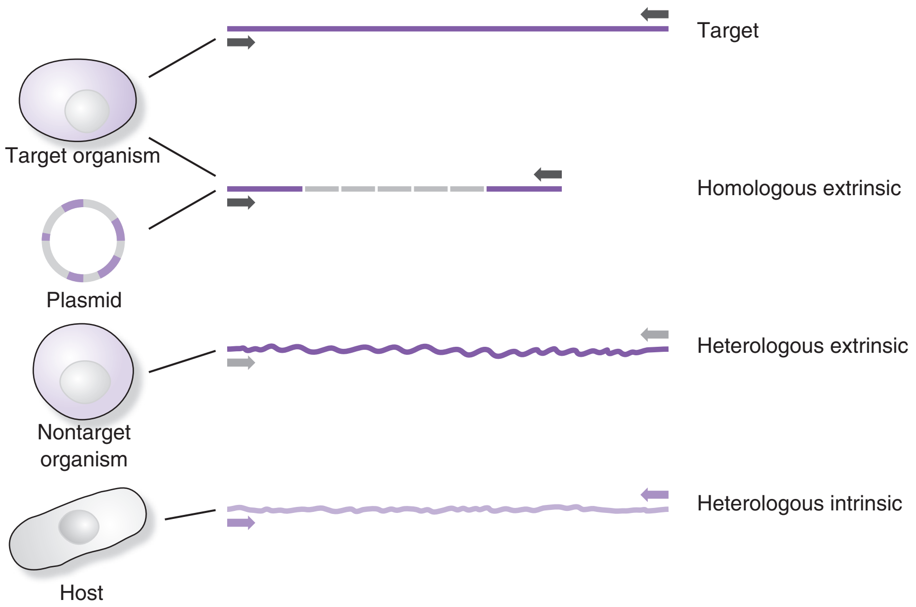


FIGURE 11.2 Amplification controls and their relationships to the molecular target. Homologous extrinsic controls are a modified version of the target that maintain the target primer binding sites. The homologous intrinsic control may be smaller, larger, or the same size as the target. Heterologous extrinsic controls are obtained from unrelated nontarget organisms and require primers different from those of the target sequences. Heterologous intrinsic controls are similar to heterologous extrinsic controls, except that heterologous intrinsic controls come from host sequences.

whereas a negative result indicates that the organism is not present (at amounts up to the detection limits of the assay). Although most false-positive test results can be eliminated by preventing carryover contamination, another source of false-positive test results that cannot be controlled in the laboratory is the presence of dead or dying microorganisms in the sample of a patient taking antimicrobial agents. In this situation, the nucleic acid–based tests will remain positive longer than culture assays and thus may appear as a **false positive.** Repeating the nucleic acid–based assay 3 to 6 weeks after antimicrobial therapy is more likely to yield a true-negative result. False-negative results may be more problematic and arise when the target organism is present but the test result is negative. There are a few reasons for false-negative results on a sample. First, the organism may be present, but the nucleic acid was degraded during collection, transport, and/or extraction. This degradation can be prevented by proper specimen handling, effective transport media, and inhibiting the activity of DNases and RNases that may be present in the sample and in the laboratory. Second, amplification procedures may be inhibited by substances present in the specimen. Hemoglobin, lactoferrin, heparin and other anticoagulants, sodium polyanethol sulfonate (anticoagulant used in blood culture media), and polyamines have been shown to inhibit nucleic acid amplification procedures.[10] Attention to nucleic acid isolation procedures and ensuring optimal purification of nucleic acid from other components of the specimen and extraction reagents will help minimize the presence and influence of inhibitors on the amplification reaction.[11] As with all clinical tests, **validation** must be performed on new molecular-based tests that are brought into the laboratory. Controls must be tested, and the sensitivity, specificity, and reproducibility of the assay must be determined using reference materials.[12] Proficiency testing of methods and competency testing of personnel should be performed regularly. The Clinical and Laboratory Standards Institute, Association for Molecular Pathology, and the Food and Drug Administration (FDA) have guidelines for molecular methods in the laboratory.

Selection of Sequence Targets for Detection of Microorganisms

Molecular methods are based on sequence hybridization or recognition using known nucleic acid sequences (primers or probes). These tests are limited by the choice of target sequences for primer or probe hybridization. The primary nucleotide sequence of many clinically important microorganisms is available from the National Center for Biological Information (NCBI) or from published literature. The specificity of molecular methods targeting these sequences depends on the primers or probes that must hybridize specifically to the chosen point in the genome of the microorganism.

Choosing a sequence target is critical for the specificity of a molecular test (Fig. 11.3). Many microorganisms share the same sequences in evolutionarily conserved genes. These sequences would not be used for detection of specific strains as they are likely to cross-react over a range of organisms. Sequences unique to the target organism are therefore selected. Some organisms such as HIV or other retroviruses have variable sequences within the same species. Such variations may be informative in, for instance, determining drug resistance or for epidemiological information; however, not all types would be detected by a single sequence. The variable sequences may be included in the probe or primer areas to differentiate between types. These type-specific probes/primers are used in a confirmatory test after an initial test using probes or primers directed to a sequence shared by all types.

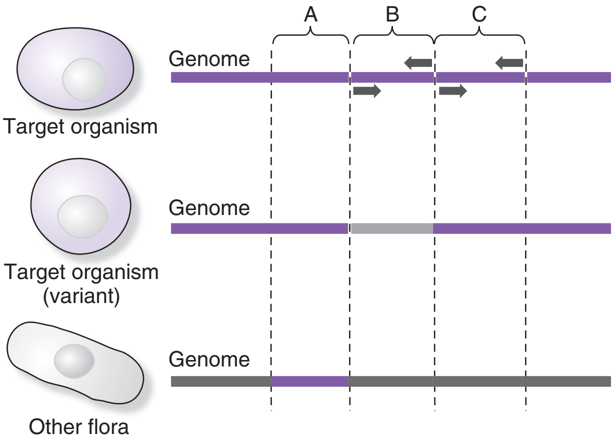


FIGURE 11.3 Selection of target sequences for a nucleic acid test. The genomes of three organisms—the test target, a variant or different type of the test target, and another nontarget organism—are depicted. Sequence region A is not specific to the target organism and is, therefore, not an acceptable area for probe or primer binding to detect the target. Sequences B and C are specific to the target. Sequence B is variable and can be used to detect and type the target, although some variants may escape detection. Sequence C will detect variants of the target organism but cannot be used for determining the type.

In addition to their strain or species specificity, target sequences must meet technical requirements for hybridization conditions. Primers should have similar annealing temperatures and yield amplicons of appropriate size. Probes must hybridize specifically under the conditions of the procedure. Sequence differences can be distinguished using sequence-specific probes or primers.

Design of probe-based amplification or detection methods, includes decisions as to the length and sequence structure of the probe, whether the probe is DNA, RNA, or protein, and how the probe is labeled. The source of the probe is also important, as probes must be replenished and perform consistently over long-term use. Probes are manufactured synthetically or biologically by cloning. Synthetic oligonucleotides may be preferred for known sequences where high specificity is required. Primer design includes the length and any modifications of the primers and type of signal generation for quantitative PCR.

Many tests currently used in molecular microbiology are supplied as commercially designed systems, including prevalidated probes and/or primers. Several of these are FDA-approved or FDA-cleared methods (http://www.fda.gov). Manufacturers of these commercial reagents specify requirements for quality assurance, including controls and assay limitations. Each system must be validated in the testing laboratory on the type of specimen used for clinical testing, including serum, plasma, cerebrospinal and other body fluids, tissue, cultured cells, and organisms. In addition to the commercial reagent sets, many professionals working in medical laboratories have developed in-house laboratory protocols (laboratory-developed tests [LDTs]) for which primers are designed based on sequence information that has been published; the reagents are bought separately, and the procedures are developed and optimized within the individual laboratory.

MOLECULAR DETECTION OF MICROORGANISMS

Molecular-based methods that have been used to detect and identify bacteria include nucleic acid-based hybridization and amplification procedures. Target detection is accomplished by a variety of methods, including agarose gel electrophoresis, amplification methods (PCR, TMA, loop-mediated isothermal amplification [LAMP]), sequencing, immunoassays, western blots, and mass spectrometry.

Real-time PCR, or quantitative PCR (qPCR), is used frequently for the detection of infectious agents because it provides a sensitive, safe closed-tube assay with quantitative information not available from conventional PCR or other "end-point" amplification methods. The quantitative capability of qPCR allows distinction of subclinical levels of infection (qualitatively positive by conventional PCR) from higher levels with pathological consequences. Furthermore, qPCR programs can be designed to provide closed-tube sequence or typing analysis by adding a melt-curve temperature program following the amplification of the target (Fig. 11.4). Like conventional PCR, qPCR is performed on nucleic acid extracted directly from clinical specimens, including viral, bacterial, and fungal pathogens.

Designing a qPCR method requires selecting a target gene unique to the specimen or specimen type for which primers and probes can be designed. The DNA-specific dye, SYBR green, can be used in place of probes if the amplicon is free of artifact, such as mis-primes or primer dimers. The probe types most often used include fluorescent energy transfer hybridization or hydrolysis probes.

The requirement for probes in addition to primers increases the complexity of the design process. Instrument software and several websites offer computer programs that automatically design primers and probes on submitted sequences. Commercial primer and probe sets are also available for purchase as reagent sets with optimized buffers and required reaction components. A variety of gene targets have been used for qPCR detection of a number of organisms. A list of examples of

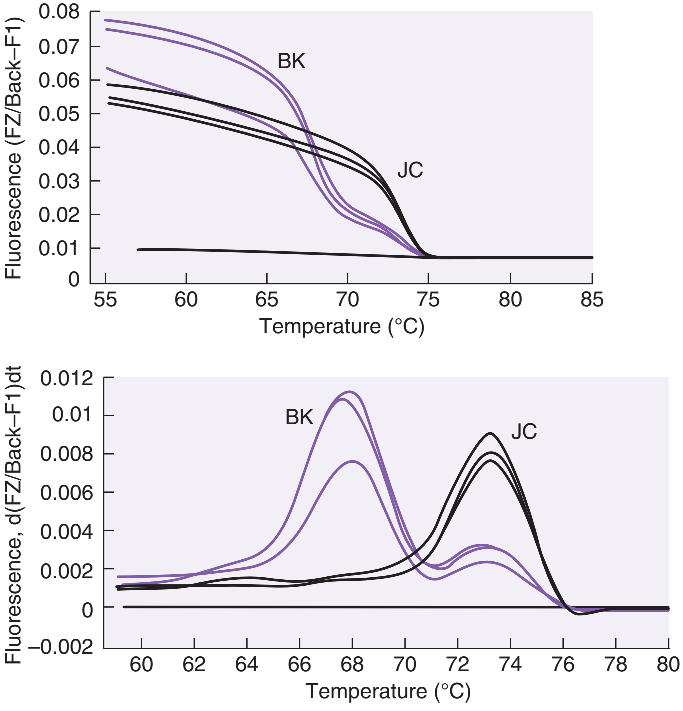


FIGURE 11.4 Melt-curve analysis of BK and JC viruses. BK and JC are differentiated from one another by differences in the T_m* of the probe specific for each viral sequence. Fluorescence from double-stranded DNA decreases with increasing temperature and DNA denaturation to single strands *(top panel)*. Instrument software will present a derivative of the fluorescence *(bottom panel)* where the T_ms (67°C to 68°C for BK and 73°C to 74°C for JC) are observed as peaks. See Color Plate 9.

targets and probes is presented in a comprehensive review by Espy et al.[13]

Similar to standard PCR, useful genes for qPCR methods include ribosomal RNA (rRNA), both 16S and 23S, and housekeeping genes such as *groEL, rpoB, recA,* and *gyrB.* The 16S rRNA is a component of the small subunit of the prokaryotic ribosome, and the 23S rRNA is a component of the large subunit of the prokaryotic ribosome. Sequencing of the DNA region encoding 16S rRNA (rDNA) is performed to determine the evolutionary and genetic relatedness of microorganisms and has driven changes in microorganism nomenclature.[14] The rDNA that encodes the rRNA consists of alternating regions of conserved sequences and sequences that vary greatly from organism to organism. The conserved sequences encode the loops of the rRNA and can be used as a target to detect all or most bacteria. Sequences that have a great amount of heterogeneity encoded in the stems of the rRNA can be used to detect a specific genus or species of bacteria. Ribosomal RNA was the original target of many bacterial molecular-based assays, but because of the instability and difficulty in analyzing RNA, current assays amplify and detect rDNA sequences and proteins.

Mass spectrometry of microbial proteins has been applied to microbiological identification and epidemiology.[15,16] In MALDI technology, proteins are converted into singly charged ions in an energy-absorbent matrix. For microbiological applications, the matrix is an acidic compound such as sinapinic acid, or a-cyano-4-hydroxycinnamic acid (CHCA) dissolved in an organic mixture of ethanol or methanol and a strong acid. The solvents penetrate cell walls and membranes, extracting the intracellular proteins. Some organisms can be spotted directly from a single colony and covered with matrix. Initial extraction in formic acid is required for reproducible identification of gram-positive organisms and fungi.[17]

Peptide databases are the central determinant in mass spec. These profiles, also called protein mass fingerprints, are maintained by instrument manufacturers and also available as open-source options. In vitro diagnostics (IVD) versions of databases of over 200 profiles are used to identify species and strains. Several factors, including culture, sample preparation, and instrument technology, influence the informativity of profiles produced.

Bacteria

Respiratory Tract Pathogens

Bacteria that cause respiratory tract disease are ubiquitous in the environment and are endemic (native to a certain region or population group) even in higher socioeconomic countries. Bacteria in the respiratory tract are easily transmitted by contact with infected respiratory secretions. Laboratory detection and identification of these organisms by nonmolecular methods often lack sensitivity and/or are time-consuming. Because of their importance in causing human disease, molecular-based assays that can detect and identify bacterial pathogens directly in respiratory specimens have been developed (see Table 11.2).

Frequent testing targets include *Bordetella, Legionella, Mycobacteria, Chlamydia,* and *Streptococcus* species. Individual IVD and analyte-specific reagent (ASR) systems have been marketed for individual testing from a variety of specimen sources. Multiplex tests are also performed for screening or speciation.

Bordetella pertussis is a pathogen of the upper respiratory tract that is the causative agent of whooping cough. The organism is endemic worldwide and is transmitted via direct contact with infected respiratory secretions. Primer and probe ASR for *B. pertussis* and *Bordetella parapertussis* detection by qPCR targeting IS*481* and IS*1001,* respectively, have been available.

Legionella pneumophila is the cause of Legionnaires' disease, an infection of the lower respiratory tract that was first described in men attending an American Legion convention in Philadelphia in 1976. Since their first identification, *Legionella* species have been found in water, both in the environment as well as in air conditioners and hot water tanks in various types of buildings. *Legionella* species infections range from asymptomatic to fatal and are the third most common cause of community-acquired pneumonias.[18] PCR tests for *Legionella* have targeted the macrophage infectivity potentiator *(mip)* gene and 16S and 5S rRNA genes.

M. tuberculosis is an important cause of respiratory tract infections causing significant levels of morbidity and mortality. The diagnosis of tuberculosis (TB) may take prolonged periods, during which time infections can spread. The genome of *M. tuberculosis* has 4.4 million base pairs (bp) with about 4,000 genes. The genomes of

TABLE 11.2 Typical Respiratory Tract Organisms Targeted by Molecular-Based Detection Methods[73,74]

Organism	Specimen Source	Gene Target	Traditional Diagnostic Methods
Mycoplasma pneumoniae	Bronchoalveolar lavage	16S rRNA 16S rDNA Species-specific protein gene P1 adhesion gene	Culture Serology
Chlamydophila pneumoniae	Respiratory Throat Atherosclerotic lesions	Cloned Pst I fragment 16S rRNA MOMP	Culture
Legionella	Deep respiratory secretions Serum Buffy coat Urine	5S rRNA *mip* gene 16S rRNA	Culture Antigen detection
Bordetella pertussis	Nasopharyngeal	IS 481 Adenylate cyclase gene Porin gene Pertussis toxin promoter region	Culture Direct fluorescent antibody
Streptococcus pneumoniae	Blood Cerebrospinal fluid Serum Sputum	DNA polymerase gene *plyA* (pneumolysin) *lytA* (autolysin) *pbp2a* (penicillin-binding protein) *pbp2b* *pspA* (pneumococcal surface protein)	Culture
Mycobacterium tuberculosis	Sputum Bronchoalveolar lavage Bronchial washings Gastric aspirates	16S rRNA	Culture

different isolates of *M. tuberculosis* do not vary to any great extent, and most variation is due to the movement of insertion elements rather than to point mutations.[19]

For many years, tuberculosis was detected from mycobacterial smears and culture. Whereas a fluorochrome stain has increased sensitivity compared with the Kinyoun and Ziehl–Neelsen stains for detecting mycobacteria directly in clinical specimens. The sensitivity of smears in general for mycobacteria varies. At least 10^4 organisms/mL are required in order to see mycobacteria in a smear, and even then, not all of those smears read as positive. Cultures for *M. tuberculosis* are more sensitive than smears and are able to detect 10^1 to 10^2 organisms/mL of specimen; however, they take time due to the slow in vitro growth of the organism. Liquid-based culture systems improved the detection rate of mycobacteria to a few days, depending on the organism load.

Nucleic acid amplification methodologies can detect *M. tuberculosis* directly in a clinical specimen with reliable sensitivity and specificity. PCR tests targeting the species-specific sequences, such as IS*6110* and 16S rRNA, allow detection of *M. tuberculosis* from fresh, frozen, or fixed tissue. PCR-positive samples are hybridized with genus-specific and species/complex-specific probes. qPCR assays have also been developed for *M. tuberculosis* detection with primers and probes targeting rRNA **internal transcribed spacer (ITS)** elements in *M. tuberculosis*.

Mycoplasma pneumoniae has been subjected to amplification techniques and other characterizing methods such as multilocus variable-number tandem-repeat (VNTR) analysis, multilocus sequence typing, and matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS).[20]

MALDI libraries of 50 to over 300 *Mycobacterium* peptide spectra are offered by at least one manufacturer. Due to the structure of their cell walls, mycobacteria require special preparation for mass spectrometry testing. The bacteria are lysed by bead beating or in boiling water (also a safety measure), extracted with ethanol, dried, and resuspended in formic acid and acetonitrile for analysis. Several studies have shown favorable comparison of MALDI-TOF with conventional methods.

Chlamydophila pneumoniae is an obligate intracellular pathogen that causes 10% of community-acquired pneumonias and has been implicated in atherosclerosis and coronary artery disease. Despite the problems with the development, implementation, and interpretation of molecular-based assays for *M. pneumoniae, L. pneumophila, B. pertussis,* and *C. pneumoniae* individually, multiplex nucleic acid amplification tests offer sensitive, specific, and rapid detection.[21,22]

Streptococcus pneumoniae is a major cause of community-acquired pneumonia and is also a common cause of bacteremia, sepsis, otitis media, and meningitis. Molecular-based tests targeting *S. pneumoniae* have attempted to detect *S. pneumoniae* in various clinical samples by targeting a variety of genes (see Table 11.2). Although PCR is specific for *S. pneumoniae,* the clinical significance of a positive PCR assay is questionable because a significant portion of the population (especially children) is colonized with the organism, and PCR cannot discern between colonization and infection.[23,24] Sequencing of 16S rRNA also does not allow enough discrimination among alpha-hemolytic *Streptococci* species because they share more than 99% sequence homology. Even MALDI-TOF identification *S. pneumoniae* is limited by database discrimination between closely related species. Improvements in sample preparation methods, the use of peak analysis, and updated databases have been applied to distinguishing these species.[25,26]

Urogenital Tract Pathogens

Neisseria gonorrhoeae and *Chlamydia trachomatis* were among the first organisms to be targeted for detection in clinical specimens by molecular methods. The molecular methods are so well characterized for these two organisms that they are used almost exclusively in the detection of the nucleic acid of *N. gonorrhoeae* and *C. trachomatis*. Other sexually transmitted bacteria are considered good targets for the development of molecular-based methods because traditional laboratory methods of detection and identification for these organisms either lack sensitivity or are time-consuming. Table 11.3 summarizes the molecular-based tests that have been described for the bacteria that cause genital tract infections.

Cultures for *N. gonorrhoeae* and *C. trachomatis* have been considered the gold standard, but nucleic acid amplification assays have the advantages of being rapid, and testing can be batched and automated, resulting in further savings for the laboratory. The first molecular-based assay available for *N. gonorrhoeae* and *C. trachomatis* was a nonamplification-based nucleic acid hybridization method that detected the rRNA with an acridinium-labeled single-stranded DNA probe. DNA probes have comparable sensitivities and specificities to culture methods. Adding enhanced signal amplification to the probe methods increased sensitivity. Numerous nucleic acid amplification assays are available that target *N. gonorrhoeae* and *C. trachomatis*. These include target amplification assays, strand displacement amplification, and transcription-mediated amplification. The nucleic acid amplification assays are performed on urethral or cervical swabs, urine, and, in some cases, on transport vials that are used to collect cervical cells for Papanicolaou smears with good sensitivity and specificity. Although molecular-based assays are the main methods for detection of *N. gonorrhoeae, C. trachomatis* cultures may still be performed in conjunction with molecular-based assays in cases of suspected child abuse. Another consideration is laboratory testing to confirm the cure of an infection. As with any infectious organism and molecular-based test, the nucleic acid is detectable in a clinical sample whether the organism is dead or alive. It is recommended that a sample should not be taken for 3 to 4 weeks after treatment to confirm treatment efficacy. Collection of samples and testing too soon after treatment will result in positive results after cultures on the same specimen have become negative.

The spirochete *T. pallidum* subspecies *pallidum* is the causative agent of syphilis, a sexually transmitted disease that results in the formation of a chancre at the

TABLE 11.3 Typical Genital Tract Organisms Targeted by Molecular-Based Detection Methods[75]

Organism	Specimen Sources	Traditional Diagnostic Methods	Gene Target
Treponema pallidum	Genital ulcers Blood Brain tissue Cerebrospinal fluid Amniotic fluid Placenta Umbilical cord Fetal tissue Serum	Serological (indirect and direct) Direct antigen detection (dark field, direct fluorescent antibody [DFA])	TpN44.5a TpN19 TpN39 p01A TpN47 16S rRNA *polA*
Mycoplasma genitalium	Urine Urethral Vaginal Cervical	Culture	MgPa (adhesion gene) rDNA gene
Mycoplasma hominis	Genital tract Amniotic fluid	Culture	16S rRNA
Ureaplasma urealyticum	Genital tract Amniotic fluid	Culture	16S rRNA Urease gene
Haemophilus ducreyi		Gram stain Culture Serological	1.1 kb target *groEL* gene Intergenic spacer between 16S and 23S rDNA p27 16S rDNA gene
Neisseria gonorrhoeae	Urine Urethral Cervical Thin preparation vials	Culture	*omp* III gene *opa* gene Cytosine DNA methyltransferase gene *cPPB* gene Site-specific recombinase gene
Chlamydia trachomatis	Urine Urethral Cervical Thin preparation vials Conjunctiva	Culture Enzyme immunoassay DFA	MOMP 16S RNA

site of inoculation (primary syphilis). If left untreated, the organism disseminates through the body, damaging tissues. The patient may progress into the other stages of disease: secondary syphilis (disseminated rash), latent syphilis (asymptomatic period), and tertiary syphilis (central nervous system and cardiovascular manifestations).

Laboratory diagnosis of syphilis is limited to serological testing, in which patients are typically screened (rapid plasma reagin [RPR] and venereal disease research laboratory [VDRL] tests) initially for the presence of antibodies against cardiolipin (a normal component of host membranes) and confirmed by testing for the presence of antibodies against *T. pallidum* by enzyme immunoassay to confirm infection. *T. pallidum* cannot be grown in vitro. Laboratories have adopted hemagglutination assays and enzyme immunoassays (EIAs) to screen patients for syphilis and fluorescent antibody absorption, particle agglutination assays, and the RPR to monitor the effectiveness of treatment and to detect reinfection.[27,28] RPR and VDRL are limited in that, when reactive, they are not specific for syphilis, and the sensitivity of these tests in very early and late syphilis is low. The serological tests that detect *T. pallidum* antibodies are limited by their inability to differentiate between current and past infections. The RPR test, though, can be used to detect reinfection because titers of anticardiolipin antibodies will decrease to nonreactive following successful treatment of the organism and increase again with reinfection.

Several PCR assays have been developed and tested for the direct detection of *T. pallidum* DNA in genital ulcers, blood, brain tissue, cerebrospinal fluid, serum, and other samples, with varying sensitivities. Amplification of the *T. pallidum* DNA polymerase I gene *(polA)* is highly accurate when tested on genital, anogenital, or oral ulcers. PCR assays for *T. pallidum* that require less than 1 mL of blood have accuracy comparable to serology.

Mycoplasma hominis, Mycoplasma genitalium, and *Ureaplasma urealyticum* cause nongonococcal urethritis. The mycoplasmas, as discussed previously for *M. pneumoniae,* are the smallest free-living, self-replicating organisms known. *M. genitalium* has the smallest genome and thus was one of the first organisms to have its genome fully sequenced.[29] *M. genitalium* culture methods were labor intensive and not widely available.

PCR assays were developed targeting the adhesion gene *(MgPa)* or the rDNA gene of *M. genitalium.* PCR was vital in establishing *M. genitalium* as an important genital tract pathogen, and laboratory-developed PCR with hybridization is used for clinical detection of *M. genitalium* as well as *M. hominis* and *U. urealyticum* detection. Assay development was at one time hindered due to the absence of a reliable gold standard for comparison and especially in the absence of clinical symptoms. Thus, the clinical significance of PCR-positive specimens was difficult to interpret. Genital tract specimens have been the target for the development of multiplex assays in which the presence of nucleic acid of multiple organisms can be determined from one specimen in a single tube. The organisms causing genital tract infections often overlap in their symptoms, or infections can be caused by the presence of multiple organisms at the same time. Since this first description of the multiplex assay, other groups have used multiplex methods to confirm the sensitivity of the assay for the targeted organisms and to examine the prevalence of these organisms in various geographic areas in different years.[30,31]

Viruses

Evidence for virus infection has been detected by testing for antibodies against the virus, by measuring the presence or absence of viral antigens, or by detecting the growth of a virus in a culture system. Although some of these methods are well established for certain viruses, they all have major disadvantages.

Antibody detection is an indirect method of diagnosis. The host immune response has to be stimulated by the virus to produce antibodies. In the case of an immunodeficient patient, the lack of antibodies due to host factors might be interpreted as a lack of the virus. Furthermore, antibodies are a retrospective indication of the infection. To interpret antibody testing with the most confidence, paired sera should be collected, with one sample collected during the acute phase of the infection and the other collected as the patient is recovering. The titers of antibodies are measured in both samples. A fourfold or greater rise in titer level from the acute sample to the convalescent sample indicates the presence of the virus during the acute stage. Detecting IgM antibodies, in particular, during an acute infection is the best evidence for the presence of a virus. But in patients in the very early stages of infection, IgM titers may be below detection limits. During this “window” period, the patient is infected and infectious.

Antigen detection testing is available in the clinical laboratory only for some viruses. Assays that measure viral antigens are available more often for respiratory syncytial virus (RSV), influenza virus, rotavirus, HSV, cytomegalovirus (CMV), and hepatitis B virus (HBV). Viral antigens are detected by enzyme immunoassays or direct immunofluorescent assays.

The classical method for detection and identification of viruses in body fluids is tissue or cell culture. Monolayers of host cells are grown in vitro, the patient’s specimen is inoculated onto the cells, and changes in the cells due to viral infection, called **cytopathic effect,** are observed microscopically by the technologist. The identity of the virus is confirmed using fluorescently labeled monoclonal antibodies. Culture has been the gold standard for many viruses, in particular adenovirus, enteroviruses, CMV, influenza, and HSV, but it is not applicable to other viruses, such as the hepatitis viruses, because these viruses do not grow well in culture systems. Another disadvantage to viral culture is the amount of time required before viral growth is detectable. Although the shell vial system, where the sample is centrifuged onto a cell monolayer and tested with antiviral antibodies, decreased detection time, several days to weeks, depending on the virus, can pass before detection of cytopathic effect.[32] Furthermore, some viruses do not produce a cytopathic effect on infecting cells, or the cytopathic effect that is produced is subtle and easily missed. In these cases, the cultures will be reported as a false-negative result. Another disadvantage of using viral cultures to detect and identify viral infections is that the specimen must be collected in the acute phase of the disease, that is, in the first 5 days of the illness, after which the amount of virus in body fluids decreases significantly and may result in false-negative cultures.

Nucleic acid amplification assays have become indispensable in the clinical virology laboratory. Molecular methods are well suited to targeting the various configurations of nucleic acids found in human pathogenic viruses (Table 11.4). Target amplification assays such as PCR, reverse transcriptase PCR (RT-PCR), quantitative PCR (qPCR), and transcription-mediated amplification (TMA) as well as signal amplification assays such as branched DNA (bDNA) amplification and hybrid capture are used in the clinical virology laboratory to diagnose or monitor viral infections. Table 11.5 summarizes the viruses for which nucleic acid amplification assays have been developed along with the type of amplification procedure, the targeted genes, and clinical utility.

Molecular-based tests for HIV, Epstein–Barr virus (EBV), human papillomavirus (HPV), HCV, and BK/JC viruses are frequently used in clinical laboratories.

Human immunodeficiency virus (HIV) is an RNA virus that makes a DNA copy of genomic RNA using its own virally encoded reverse transcriptase. There are two types of HIV: HIV type 1, or HIV-1, and HIV type 2, or HIV-2. HIV-2 is a minor isolate found mainly in West Africa and is less pathogenic than HIV-1, causing more latent infections. New strains continue to evolve.

HIV rapidly mutates and recombines so that multiple groups then divide into "clades," or subtypes, of the virus that are found in different locations in the world. There are four HIV subtypes, M (Major), O (Outlier), N (non-M and Non-O), and P. HIV group M causes 95% of the infections due to HIV around the world. Group M is further divided into eight clades (A, B, C, D, F, G, H, and J). Group M, clade B, is found most often in

TABLE 11.4 Genomes of Human Viruses

Double-Stranded DNA Viruses	Adenovirus
	BK virus
	Cytomegalovirus
	Epstein–Barr virus
	Hepatitis B virus
	Herpes simplex virus 1
	Herpes simplex virus 2
	Human papillomavirus
	JC virus
	Molluscum virus
	Rotavirus
	Vaccinia virus
	Varicella-zoster virus
Single-Stranded DNA Virus	Parvovirus
Double-Stranded RNA Virus	Rotavirus
Single-Stranded RNA Viruses	Colorado tick fever virus
	Coronavirus*
	Coxsackie virus*
	Dengue virus*
	Ebola virus[†]
	Echovirus*
	Hepatitis A virus*
	Hepatitis C virus
	Hepatitis D virus
	Hepatitis E virus
	Human T-cell leukemia virus[‡]
	Human immunodeficiency virus[‡]
	Influenza virus
	Measles virus[†]
	Mumps virus[†]
	Norwalk virus, norovirus
	Parainfluenza virus
	Poliovirus*
	Rabies virus
	Respiratory syncytial virus
	Rhinovirus*
	Rubella virus*
	Yellow fever virus*

*Positive RNA; directly translated.

[†] Negative RNA; complementary to the translated strand.

[‡] Retroviral replication requires a DNA intermediate.

TABLE 11.5 Nucleic Acid Amplification (NAA) Tests for Viruses

Virus	NAA Methodology	Amplified Target	Dynamic Range/ Sensitivity	Clinical Utility
Human immunodeficiency virus	PCR RT-PCR NASBA bDNA	*gag* gene; 155 bp (HIV-1 group M [subtypes A-H], not HIV-2 or HIV-1 group O) *gag* (similar to Amplicor) HIV-1 groups M, O, and N *pol;* subtypes of group M (subtypes A–G, but not group O)	400–750,000 copies/mL (standard) 50–100,000 copies/mL (ultrasensitive) 176,000–3,470,000 copies/mL 75–500,000 copies/mL	Viral quantitation Disease prognosis Treatment monitoring
Cytomegalovirus (CMV)	Hybrid capture PCR NASBA qPCR Major capsid protein	Immediate-early antigen 1 Major immediate-early antigen Glycoproteins B and H Detect HSV when asymptomatic or when cultures are negative	1,400–600,000 copies/mL 400–50,000 copies/mL	Detect CMV DNA in organ transplant and AIDS patients and congenitally infected infants Viral load determinations
Herpes simplex virus-1 and -2 (HSV)	PCR qPCR	Thymidine kinase DNA polymerase DNA-binding protein Glycoproteins gb, gc, gd, and gg		Diagnosis of HSV encephalitis and neonatal infections
Epstein–Barr virus (EBV)	PCR qPCR	EBNA1 LMP-1	100–10^{10} copies/mL	EBV-associated malignancies Detect EBV in asymptomatic immunocompromised hosts
Human papillomavirus (HPV)	Hybrid capture PCR qPCR	L1 or E1 open reading frames	10^5 copies/mL	Detection of HPV in endocervical swabs Differentiation of low-risk and high-risk types Monitoring women with abnormal Pap smears
Hepatitis B virus	PCR bDNA Hybrid capture		1,000–40,000 copies/mL 0.7–5,000 meq/mL 142,000–1,700,000,000 copies/mL (standard) 4,700–56,000,000 copies/mL (ultrasensitive)	Prognosis and monitoring of antiviral treatment response
Parvovirus B19	PCR			Diagnosis of infections
Respiratory syncytial virus (RSV)	RT-PCR	Fusion glycoprotein (F) gene Nucleoprotein (N) gene		Detection of RSV Differentiate between subgroups A and B

Continued on following page

TABLE 11.5 Nucleic Acid Amplification (NAA) Tests for Viruses *(Continued)*

Virus	NAA Methodology	Amplified Target	Dynamic Range/ Sensitivity	Clinical Utility
Parainfluenza viruses	RT-PCR	Hemagglutinin-neuraminidase conserved regions 5′ noncoding region of F gene		Epidemiology
Influenza viruses	RT-PCR	Conserved matrix (M) genes (influenza A and B) Nucleocapsid protein (influenza A) NS1 gene (influenza B)		Diagnose infections Characterize isolates Can be type- and subtype-specific
Metapneumovirus	RT-PCR	Fusion (F) gene RNA polymerase (L) gene		
Coronavirus	RT-PCR	RNA polymerase gene Nucleoprotein gene		Used to detect and characterize the SARS virus
Norwalk virus	RT-PCR	RNA polymerase gene		
Rotavirus	RT-PCR	VP7 gene VP4 gene		
Hepatitis C virus	RT-PCR bDNA	5′ untranslated region (UTR) 5′ UTR and core protein gene	600–800,000 IU/mL 3,200–40,000,000 copies/mL	
West Nile virus	RT-PCR NASBA	Variety of gene targets based on genome of type strain NY99 0.01 PFU	0.1–1 PFU	Used for diagnosis and surveillance
Rubella virus	RT-PCR	Surface glycoprotein, E1, gene	Variable; sensitivity of 3–10 copies is best	Primarily for fetal diagnosis Used for diagnosis when serum is not available May be used to confirm positive serological results
Mumps virus	RT-PCR	Hemagglutinin, neuraminidase, P, SH, and F genes		Differentiate strains
Measles virus	RT-PCR	M, H, F, N		When culture is not practical or genotyping is required for diagnosis of MIBE or SSPE Differentiation of vaccine and wild-type strains

TABLE 11.5 Nucleic Acid Amplification (NAA) Tests for Viruses *(Continued)*

Virus	NAA Methodology	Amplified Target	Dynamic Range/ Sensitivity	Clinical Utility
Enteroviruses (group A and B Coxsackieviruses, echoviruses, and others)	RT-PCR	Conserved 5′ nontranslated region		Performed on CSF to rule out enteroviral meningitis
BK virus (polyomavirus)	PCR	Large T protein		Diagnosis of BK virus nephritis

MIBE, Measles inclusion body encephalitis; *PFU,* plaque-forming units; *SSPE,* subacute sclerosing panencephalitis.

the United States and Europe. Group O HIV is found primarily in West Africa, and group N is found in Cameroon. To infect host cells, the HIV surface molecules gp120 and gp41 interact with CD4, a molecule that is expressed primarily on the surface of helper T lymphocytes but is also found on macrophages, dendritic cells, and other antigen-presenting cells. Chemokine receptors, in particular CCR5, on dendritic/Langerhans' cells and macrophages/monocytes, and CXCR4 on CD4+ T cells form a complex with CD4 on the cell surface and also engage gp120. After attachment to host cells via CD4-gp120 binding, the virus enters the cell, where reverse transcriptase makes cDNA from viral RNA. The cDNA integrates into the host DNA, where it either persists in a stage of latency as a provirus or is replicated actively. Transcription and translation of viral peptides, as well as production of viral RNA, are performed by cellular components under the direction of virally encoded regulatory proteins (i.e., *tat, rev, nef,* and *vpr*).

HIV infection is identified as antibodies specific for HIV in an EIA and by confirming the specificity of detected antibodies for HIV products in a western blot or a qualitative RNA probe assay. The antigens used in the western blot tests may differ depending on the country of origin. For infants who have maternal IgG and for patients suspected of incubating HIV in whom antibody tests are negative, antigen detection tests measure the amount of HIV p24 antigen. Quantitative nucleic acid amplification assays are performed after an HIV diagnosis to determine how actively the virus is replicating (viral load), when to start antiretroviral therapy, and when to monitor the efficacy of treatment.

The amount of HIV or **viral load** is used as a marker for disease prognosis as well as to track the efficacy of antiretroviral therapy. The goal of antiretroviral therapy is a viral load below 50 copies/mL of blood, where it is undetectable by most methods. Patients who maintain viral levels at fewer than 10,000 copies/mL in the early stages of the infection are at decreased risk of progression to acute immunodeficiency syndrome (AIDS).[33]

Patients who are effectively treated with antiretroviral therapy will have a significant reduction in viral load 1 week after the initiation of therapy. The lack of a significant decrease in viral load during this time indicates the lack of efficacy. Highly active antiretroviral therapy (HAART), consisting of two reverse transcriptase inhibitors combined with a protease inhibitor or a nonnucleoside reverse transcriptase inhibitor, reduced viral loads below the detection limits of even ultrasensitive assays.[34] In patients receiving HAART, 2 $\log_{10}$ decreases in viral load have been documented. Viral load testing is performed in conjunction with determining CD4 counts. In general, but not always, viral load and CD4 counts are inversely proportional; that is, the higher the viral load, the lower the CD4 count.

HIV in patient samples can also be detected by PCR of integrated DNA, nucleic acid sequence–based amplification (NASBA), and bDNA (see Table 11.5). Table 11.6 compares the advantages and disadvantages of each of these assays for determining HIV viral load. Viral load should be determined before therapy is started, 2 to 8 weeks after therapy initiation to see the initial response, and then every 3 to 4 months to assess therapeutic effectiveness.

TABLE 11.6 Advantages and Disadvantages of the Nucleic Acid Amplification Methods for HIV Viral Loads

Test	Advantages	Disadvantages
bDNA	High throughput Broad dynamic range Applicable for group M subtypes A–G	No internal control False-positive results reported
Amplicor RT-PCR	Internal control Good specificity	Limited dynamic range
NASBA	Broad dynamic range Performed on many specimen types and volumes	Does not detect all non-B subtypes

TABLE 11.7 Test Performance Features for Viral Load Measurement

Characteristic	Description
Sensitivity	Lowest level detected at least 95% of the time
Accuracy	Closeness of measured value to a standard or known value
Precision	Reproducibility of independently determined test results
Specificity	Negative samples are always negative, and positive results are true positives
Linearity	A serial dilution of standard curve closely approximates a straight line
Flexibility	Accuracy of measurement of virus regardless of sequence variations

Results among methods are increasingly comparable, such as viral loads determined by bDNA and RT-qPCR.[35] Even so, it is still recommended that the same method to determine viral loads is used when monitoring patients over time. Results are expressed consistently as integers (copies/mL or IU/mL), log-transformed results (log IU/mL), scientific notation ($a \times 10^b$ IU/mL), or a combination of these. CDC guidelines recommend reporting in integers and log-transformed copies. Quantitative HIV-1 RNA testing in plasma has been the standard for monitoring drug therapy and HIV disease progression. The Multicenter AIDS Cohort Study, an ongoing project monitoring the clinical histories of treated and untreated HIV-infected men, has described laboratory measurements of viral load and clinical disease. The U.S. Department of Health and Human Services (DHHS) recommends an objective of maximal suppression of viral replication down to undetectable levels by sensitive analysis.

The first commercially produced tests could detect viral loads down to 400 to 500 copies/mL plasma; however, suppression to fewer than 20 copies/mL plasma was associated with longer response to therapy than suppression below 500 copies/mL, which emphasizes the importance of a highly sensitive assay with a very low limit of detection for optimal treatment strategy and patient care. In addition to sensitivity, all test methods, including HIV tests, should have certain characteristics (Table 11.7). Accuracy is an important requirement for viral load testing. It is established by the calibration of assays to a common standard. Quantitative tests must also have accuracy that is a true measure of the viral level over a range of values. The viral load levels established by the DHHS and the International AIDS Society for the initiation of therapy must be consistently identified in independent laboratories as accurately as possible. Quantitative PCR methods offer linear measurement over a wider range than other methods, which precludes the requirement for dilution of high-titer specimens before analysis. The precision or reproducibility of the test used is also important for establishing statistically significant differences in the viral load over a serial testing period. The DHHS defines a minimally significant change as a threefold increase or decrease in viral load/mL plasma. And the high specificity of a test gives confidence that a positive result is truly positive. All subtypes of virus should be detected with equal efficiency to avoid under- or overestimating viral loads of certain subgroups.

HIV RNA in the same patient will not change much over time (approximately 0.3 $\log_{10}$ unit) as long as the patient is clinically stable and antiretroviral therapy has not begun or changed. In order to be clinically relevant, viral load changes from one determination to another must vary by threefold (0.5 $\log_{10}$ unit). HIV-positive patients may experience transient increases in viral loads when they have other infections or receive vaccinations, but levels will return to baseline within a month. Proficiency testing is available from the College of American Pathologists (CAP; Northfield, IL) and the CDC. The World Health Organization (WHO) has an HIV-1 RNA reference standard.

Error-prone replication of HIV by its reverse transcriptase and recombination between co-infecting strains generates new HIV sequences, which can affect the antiviral drugs, including protease inhibitors, nucleoside analogs, reverse transcriptase inhibitors, and inhibitors of viral integration. Genotyping is used to monitor the development of this antiretroviral drug resistance. Most of the antiretroviral drugs target the reverse transcriptase and protease enzymes, so these are the genes that are most often examined in genotyping procedures.

To perform genotyping, viral RNA is extracted, and PCR is used to amplify the whole protease gene and part of the reverse transcriptase gene. The products are analyzed for the presence of mutations by sequencing, hybridization onto high-density microarrays, or reverse hybridization. Sequencing is performed most often, and there are commercially available kits for Sanger sequencing. Using these genotyping methods for resistance monitoring, however, is a high-cost and low-throughput method. Next-generation sequencing (NGS) along with simpler sample collection and storage matrices (e.g., dried blood spots[36]) will facilitate and broaden resistance monitoring as well.[37]

To evaluate resistance, viral sequences are compared with a reference sequence to identify the mutations present. After the mutations have been identified, their significance in terms of the impact on antiretroviral therapy is assessed, and this is generally accomplished through computer algorithms. Resistance mutations have been well characterized for individual agents, but HIV-positive patients are more often on cocktails of drugs rather than one drug. Therefore, the impact of a mutation on multiple drugs is also considered. In addition, if a single virus has multiple mutations, the interpretation of more than one mutation in the context of the others and with respect to multiple drugs becomes more complex. Computer algorithms are used to analyze genotypes, taking into account primary and secondary mutations, cross-resistance, and the interactions between mutations that can affect resistance.[38]

Mutations found by genotyping are generally divided into two groups: **primary resistance mutations** and **secondary resistance mutations.** Primary resistance mutations are those that are specific for a particular drug, reduce the susceptibility of the virus to that drug, and appear in the viral genome shortly after treatment with that agent has begun. The mutated enzyme is generally not as active as the normal enzyme, so viral replication is decreased, but it still occurs. As treatment with the drug continues, secondary or compensatory mutations occur that try to recover the ability of the virus to replicate at a normal rate. The secondary mutations do not affect the susceptibility of the virus to the drug, but rather help the virus replicate in the presence of the drug when one of its replication enzymes is not 100% functional. Once a resistance genotype has been identified, the drug therapy of the patient should be changed as soon as possible to avoid the development of secondary mutations in the virus.

The results of genotyping procedures are reported by listing the mutations that have been identified in the protease and reverse transcriptase genes and the impact those mutations will have on each drug: no evidence of resistance, possible resistance, resistance, or insufficient evidence. The mutations are indicated by reporting a change in the amino acid that is coded by the changed codon, where the wild-type amino acid is written, followed by the position of the codon that is changed, followed by the new amino acid. For example, a mutation in codon 184 of the reverse transcriptase gene from ATG to GTG results in an amino acid change from methionine to valine, or M184V. This particular mutation makes the virus resistant to the cytidine analog, lamivudine. Another mutation, Q151M, located at the dNTP-binding site of reverse transcriptase, confers resistance to reverse transcriptase inhibitors while maintaining sensitivity to lamivudine.

As with all other molecular-based assays, HIV genotyping procedures should employ adequate quality and contamination controls. The sensitivity of the methods for detecting a minority of virions that contain mutations in the midst of a majority of wild-type virions is an important consideration. The sensitivity of the automated sequencing methods has been reported to be 20%, and NGS has a reported sensitivity of less than 1%.[39] Proficiency testing is provided by the CAP, and independent control materials for use in genotyping assays are commercially available.

Herpes viruses are frequent sources of human infection. At least 25 viruses comprise the family *Herpesviridae*. Several herpes virus types frequently infect humans, including herpes simplex virus (HSV), CMV, EBV, and varicella-zoster virus (VZV). In addition to causing overt disease, herpes viruses may remain silent and be reactivated many years after infection.

HSV has a relatively large viral genome encoding at least 80 protein products. About half of these proteins are involved in the interaction with the host cell or immune system. The other half control viral structure and replication. There are two types of HSV, HSV type 1 (HSV-1) and HSV type 2 (HSV-2). HSV-1 mainly causes cold sores or fever blisters; HSV-2 mainly causes genital sores but can also cause mouth sores. HSV tests are usually done for genital sores, although body fluids, including blood, urine, tears, amniotic fluid, lavages, and spinal fluid, may also be tested.

HSV was detected by viral culture, antigen, and antibody tests before the application of PCR. Viral cultures sometimes had to be performed more than once because the virus might not be detected at all times of infection. The HSV antigen test was performed on a slide smear of material from the sore. Viral antigens are detected with labeled antibodies. The antibody test was performed on blood to detect antibodies to the viral antigens. Antibodies may not be detectable 2 weeks to 3 months after initial infection; however, once the infection occurs, antibodies remain in the person for life. Western blot tests could distinguish HSV-1 and HSV-2 using type-specific antigens to detect the corresponding type-specific antibodies. Type-specific testing is important for prognosis and patient counseling.

PCR tests also offer the advantage of distinguishing HSV-1 from HSV-2 directly without culturing. Methods include standard PCR and quantitative PCR targeting type-specific HSV genes (Table 11.5). One of the most common human viruses, EBV is a member of the herpes virus family. Most people become infected with EBV sometime during their lives. Children can become infected with EBV once they lose maternal antibody protection, and most adults in the United States between ages 35 and 40 have likely been infected with EBV. The first infection during adolescence or young adulthood causes infectious mononucleosis.

EBV establishes a permanent dormant infection in cells of the throat, blood, and immune system. Latent infection is characterized by the presence of EBV-encoded RNA (EBER). The virus can reactivate and is commonly found in the saliva of infected persons. EBV infection is associated with a variety of benign and malignant lesions, including Hodgkin disease, non-Hodgkin lymphoma, Burkitt lymphoma, gastric carcinoma, and nasopharyngeal carcinoma. Although EBV is present in these conditions, it is likely not the sole cause of disease.

Laboratory testing for EBV infection has been performed by immunohistochemistry for EBV proteins in tissue and testing serum for antibodies to EBV-associated antigens, including the viral capsid antigen, the early antigen, and the EBV nuclear antigen (EBNA). Confirmation may be done by differentiation of IgG and IgM subclasses to the viral capsid antigen. EBER transcripts are primary targets for tissue analysis. Repeated or unique sequences in EBV DNA are targets of in situ hybridization; however, the sensitivity of DNA probes is lower than those for EBER, due to the abundant presence of the EBER transcripts.

Southern blot analysis of EBV DNA, first described in 1986,[40] was based on variable numbers of terminal repeat sequences at the ends of each EBV DNA molecule. Each infecting genome can contain up to 20 terminal repeat sequences. When EBV DNA is cut with *Bam*H1, the resulting fragments vary in size, depending on the number of repeat sequences. The fragments were visualized with a probe complementary to the variable repeat.

Amplification methods are now used for detecting EBV in blood, body fluid, or tissue samples. EBV DNA can be amplified using primers complementary to conserved EBV sequences, and strain typing can be achieved by amplification of polymorphic regions of the viral genome. Quantitative PCR is used to determine EBV viral load in blood using a reference internal control. A reported quantitative range of analysis is 750 to 10^6 IU/mL.[41] As with other such tests, the virus may still be present below the level of detection.

Another member of the herpes virus family is CMV. Most people have been infected with CMV without obvious illness. Like other herpes viruses, CMV can go dormant and reactivate. If symptoms occur, they are mild, except in immunocompromised individuals. The virus is shed in blood, urine, saliva, and other body fluids but dies quickly outside of the host.

Molecular detection of CMV is performed on cell-free plasma or other fluids. DNA is extracted and amplified by PCR using primers targeting the CMV polymerase (UL54) or glycoprotein B (gB) gene in regions that do not share sequence homology with other herpes viruses. Internal controls are included to avoid false-negative interpretation due to PCR inhibition. The level of detection of standard laboratory assays is such that no PCR product will be produced from normal plasma, even from people previously exposed to CMV. Automated amplification and detection of CMV has an analytical sensitivity as low as one viral copy per microliter. Quantitative PCR can detect down to 40 genome copies of virus per microliter of blood. PCR assays may be affected by the presence of mutations in the target genes that interfere with primer hybridization. These mutations may also confer resistance to antiviral therapy.

VZV (HHV3) is a herpes virus that infects younger and older humans. VZV is the causative agent of chickenpox and shingles (also called herpes zoster). The virus has a large genome containing over 70 open reading frames (ORFs). Polymorphisms in the ORF are used to determine the strain variations and genotype of viruses from different geographical areas. An attenuated live virus, VZV Oka, developed by selection in cell cultures, is the parental strain source of a VZV vaccine, which contains a mixture of genotypically distinct viral variants.

VZV is neurotropic, remaining latent in nerve cells and possibly reactivating years after primary infections to result in shingles, or palsy. It is found in more than 90% of young people living in temperate climates. There is no animal reservoir for VZV. Clinical tests include enzyme-linked immunosorbent assays and PCR methods for detection of antibodies to VZV in serum. Qualitative PCR methods to detect VZV include PCR with hybridization or direct gel electrophoresis. Quantitative PCR kits include internal controls to detect amplification inhibition.

HCV is an enveloped, single-stranded RNA virus of the Flaviviridae viral family. HCV causes viral hepatitis and cirrhosis and is also associated with causing hepatocellular carcinoma. The virus is transmitted parenterally like HIV. Acute infections are often asymptomatic and are rarely associated with jaundice; thus, patients with acute HCV infections are usually not detected at this stage. The development of chronic infections is due to the antigenic envelope proteins that elicit the production of antibodies. These proteins are encoded by hypervariable regions much like antibody genes themselves, resulting in extensive variation in the envelope proteins and the escape of the virus from antibodies.

The initial approaches to HCV analysis are similar to those for HIV. Serology is used to detect the presence of antibodies against HCV. If the patient has HCV antibodies, the specificity of the antibodies for HCV antigens is measured by western blot, where the presence of antibodies with multiple HCV specificities confirms the diagnosis.

A variety of nucleic acid amplification assays for the qualitative detection and quantitation of HCV are available, including RT-PCR, transcription-mediated amplification, and branched DNA. The qualitative HCV RNA assays are performed on patients with positive HCV antibody results to confirm active infection or on immunocompromised patients (who are often co-infected with HIV) in whom HCV infection is suspected but antibody tests are negative. The quantitative HCV RNA assays are used as for HIV to determine the viral load and to monitor viral replication in response to antiviral therapy. The viral load and the HCV genotype are used to determine the therapeutic protocol, both type of drug(s) as well as duration.

Six genotypes (1a/1b, 2, 3, 4, 5, and 6) and more than 50 subtypes of HCV have been identified. The genotype of HCV present in a given patient determines the treatment protocol used on that patient, as particular genotypes are associated with certain antiviral resistance patterns. The HCV genotype is determined by analyzing the core and/or 5′ untranslated regions of the genome. Laboratory methods available for HCV genotyping are PCR with restriction fragment length polymorphism (RFLP) analysis and reverse hybridization and direct DNA sequencing. A PCR with melt-curve method has also been described.[42]

It has been observed that patients who have a 2 $\log_{10}$ decrease in HCV RNA 12 weeks after treatment begins have a 65% chance of responding, defined by the lack of detection of HCV RNA in qualitative assays where the detection limit is 50 to 100 copies of virus/mL of plasma. Patients who do not have a 2 $\log_{10}$ decrease in HCV RNA 12 weeks after treatment begins have less chance of responding. Determining the genotype of the virus is also critical to predicting treatment outcomes; for example, genotypes 2 and 3 will respond better to particular treatments than genotype 1.

HPV is a double-stranded DNA virus recognized as oncogenic in several human cancers. There are over 200 HPV types based on sequences of the viral genome compared with known HPV genomes. Five evolutionary HPV genotype groups (α, β, γ, μ, and ν) have been defined. The largest, the α group, contains 64 types that mainly infect mucosal epithelia in the anogenital tract and include the high risk (HR) types (HPVs 16, 18, 31, 33, 35, 39, 45, 51, 56, 58, 59, 68, 73, 82) that have been classified as oncogenic and are found to cause anogenital cancers. The next largest group is the β-group HPVs that contains more than 50 characterized types that mainly infect cutaneous epithelia. The β group and ultraviolet (UV) radiation exposure are associated with nonmelanoma squamous cell carcinomas, a common human cancer. HPVs of the remaining three groups (γ, μ, and ν) normally cause only benign disease.[43] HPV HR types are responsible for over 95% of cervical squamous carcinoma. The presence of high-risk HPV DNA in conjunction with an equivocal or ambiguous cytology result (atypical squamous cells of unknown significance [ASC-US]) indicates an increased risk for having an underlying cervical neoplasm. Persistent infection with high-risk HPV may be the main risk factor for the development of high-grade cervical neoplasia and cancer. Women with normal cervical cytology who are negative for the high-risk HPV types are at low risk for having or developing cervical precancerous lesions.

It is difficult to culture HPV in the laboratory, so detection relies on molecular testing. Two approaches to molecular detection of HPV are hybridization methods and amplification methods. The latter methods include target amplification, such as PCR, and signal amplification, such as hybrid capture. Hybridization tests can detect a minimum of 4,000 to 5,000 viral genomes. These tests also have the capacity for subtyping. HPV expresses its early genes, E1 through E8, shortly after host infection. The E6 and E7 genes are overexpressed after integration of oncogenic genotypes of the HPV genome in the host genome. E6 and E7 gene products are involved in the cellular transformation to cervical cancer. Measurement of E6 and E7 mRNA expression in specific cell types identified by flow cytometry is a specific indication of cellular transformation in the small percentage of HPV DNA infected cells that are precancerous. The PCR genotyping detects 37 high- and low-risk HPV types by hybridization of PCR products to immobilized probes for each of the mutations. Direct sequencing may also be used to detect more HPV types.

Methods may differ, not only in the sensitivity to target viruses but with the ability to detect concurrent infection with different HPV types. In all cases, patient management decisions reflect patients' overall cytology history and other risk factors in addition to the presence or absence of high-risk HPV types.

Respiratory viral infections are a major cause of hospitalizations in young children and elderly people in the United States. Over a dozen respiratory pathogens (viral and bacterial) are commonly encountered in the medical laboratory. Treatment and control of the spread of infection will depend on which of these are infecting a patient. For viral infections, molecular methods have been designed to detect multiple species in single analyses. For example, a real-time PCR method can detect influenza A, influenza B, and RSV. Influenza A subtyping by melt-curve analysis has been used as a screening method to detect the 2009 influenza A (H1N1) virus from nasal swabs.

Advanced Concepts

Amplification methods that target the HPV L1 gene may not detect virus that has integrated into the host DNA because that gene region may be lost, causing false-negative results. Coinfection with two or more HPV types will lower overall PCR efficiency for each type or selectively amplify only one type, also resulting in false-negative results.

RSV is detected using several assays. An ASR for RSV A + B RNA uses NASBA technology. A bead array technology can simultaneously detect RSV A and B, influenza A, nonspecific influenza A, H1, H3, influenza B, parainfluenza 1, parainfluenza 2, parainfluenza 3, metapneumovirus, rhinovirus, and adenovirus. The test includes an MS-2 bacteriophage internal control and a lambda bacteriophage positive control. It aided in the detection of 2009 influenza A/HIN1 (swine flu) but could not identify the hemagglutinin gene of the 2009 influenza A/H1N1 directly.[44] Seasonal and H1N1 flu viruses develop resistance to antiviral agents through sequence alterations.[45] These mutations can be detected by direct sequencing. Digital PCR methods have also been developed to detect the frequently occurring alteration H275Y in the neuraminidase gene (H274Y in N2 nomenclature).[46]

BK and JC viruses are human polyomaviruses, the primate counterpart of which is SV40. BK, JC, and SV40 share sequence and antigenic homology. They are double-stranded DNA viruses with 5,000-bp genomes encoding 5 transcripts. Most polyomavirus infections are subclinical in adults. Infected children develop respiratory symptoms, and some have cystitis. Several reagent sets are available for detection of BK or JC viruses in the clinical laboratory, including quantitative PCR tests.[47]

Mass Spectrometry

Application of mass spectrometry to bacteriology has been extended to virology to a limited degree. Viruses have low concentrations of high-molecular-weight proteins that are more difficult to assess by peptide profiles. Confounding viral detection is cell debris from the cultures in which the viruses were cultivated. MALDI analysis of PCR-amplified viral nucleic acid has been successfully applied to clinical detection of HSV, HCV, HPV, enteroviruses, and respiratory viruses.[48] Multiplex PCR with MALDI (PCR-mass assay) has been shown to identify multiple viruses or multiple viral subtypes in a single test.[49]

PCR-based genotyping of frequently rearranged viruses such as influenza where recombinant viruses arise from coinfection with distinct viral stains has proven more difficult. PCR amplification in these cases becomes problematic as primer-binding sights are lost or changed in the recombinant strains. These reassorted flu virus strains are also difficult to detect by immunological methods as the antigenic determinants become altered. Characterization of the type of virus is important for early intervention and prevention of infectious spread. For this application, mass spectrometry has been shown to successfully genotype viruses using whole-virus protein digests.[50] Human influenza and parainfluenza viruses can be subtyped and lineages traced by this method.

Mycology

Fungi are among the most ubiquitous microorganisms causing clinical infection. Traditional methods of identification by phenotype have become more difficult with the expanding diversity of organisms. Molecular methods, particularly sequencing and PCR, have allowed for the detection and typing of fungi with greater sensitivity, specificity, and speed.

Fungi are important causes of human disease, especially in immunocompromised patients. In addition, laboratory-acquired infections from fungi are a major risk for laboratory personnel. Fungal infections are most often diagnosed by direct staining methods and isolation of the causative agent in culture. As for other organisms, traditional smears and cultures are affected by sensitivity, organism viability, and the length of time required for the organism to grow. Despite these problems, direct smears and isolation of fungi are still the major methods for detecting fungi in clinical samples.

Fungi growing in culture are typically identified by their microscopic and macroscopic morphologies or by using fluorescent antibodies. For some fungi, such as *Histoplasma, Blastomyces, Coccidioides,* and *Cryptococcus neoformans,* gene probes have been developed to confirm the identity of the organism growing in culture.[51] Using DNA probes for these organisms is faster and less hazardous than determining the microscopic morphology.

Broad-range PCR and subsequent analysis are also used for clinical analysis of fungi. In these approaches, primers anneal to DNA sequences that are common to most of the clinically relevant fungi, such as *Candida, Aspergillus, Rhizopus* and other Zygomycetes, and *Histoplasma* and other dimorphic fungi. Once the sequences are amplified, hybridization to species-specific probes or sequencing is used to identify the fungus to the genus or genus and species level. In-house-developed real-time PCR methods are also used for direct identification of *Aspergillus* and *Pneumocystis carinii*. *Blastomycetes*, *Coccidioides immitis,* and *Histoplasma capsulatum* are detected with direct probe hybridization assays. The probe methods can also be multiplexed. PFGE is used in laboratory-developed methods for molecular typing of yeasts.

Molds can be typed by PCR and sequencing of ITS regions or 28S rRNA. ITS sequences of DNA extracted from 24 to 48 mold cultures are compared with those of reference strains. This method is used in industrial and environmental settings to positively identify *Paecilomyces* species, which are saprophytic filamentous fungi found in soil and as air and water contaminants. *Paecilomyces lilacinus* and *Paecilomyces variotii* are increasing causes of opportunistic human infections generally associated with the use of immunosuppression therapy, saline breast implants, or ocular surgery. Although these two species present differences in their in vitro susceptibility to antifungal agents, requiring identification before treatment, molecular testing in the clinical laboratory is not well developed.

MALDI-TOF application to detection of fungi has been limited by their biological complexity and different fungal growth phases. Furthermore, their thick cell walls require chemical and physical disruption methods. Cell walls are extracted with trifluoroacetic acid, or formic acid and acetonitrile, possibly followed by physical disruption with beads.

Mass spectrometry methods for identification of *S. cerevisiae* were first developed in 2001.[52] Reproducible spectra have since been observed for other yeasts including *Candida* and *Cryptococcus*. MALDI-TOF may also offer a reliable, rapid approach to testing for fungal sepsis.[53]

Parasites

Parasites are typically detected and identified by morphology directly in clinical specimens, a method of diagnosis subject to false-negative results in cases of low organism concentrations and depending greatly on appropriately trained personnel. Molecular-based testing for parasites has been limited mainly because parasites are not a major cause of disease in developed countries. Increasingly, however, travelers from parasite-endemic countries bring parasites to developed countries and serve as a reservoir for transmission. In the laboratory, nucleic acid template preparation from oocysts and spores of protozoan parasites is complicated by the nature of the organism and its resistance to disruption and lysis. Parasites found in complex matrices such as stool samples present a further difficulty due to inhibition of PCR and other enzymatic assays. However, expertise in identifying parasites by morphology is declining, demanding alternative methods of surveillance and detection of these organisms.

Recognition of these issues has made the development of molecular-based assays for parasite detection and identification more practical. Nucleic acid isolation for molecular testing has been addressed by using several methods, including extraction and extraction-free, filter-based protocols for preparation of DNA templates from oocysts and microsporidia spores. PCR methods are modified to accommodate sequence content, for example, low GC content in malaria.

PCR assays have been developed for the detection of Trypanosomes, Plasmodia, Toxoplasma, Entamoebae, and Cryptosporidium in hosts and water sources. Multiplex qPCR methods can also measure infection rates and organism loads in hosts and nonhuman vectors (carriers of infection).[54]

In-house-developed real-time PCR methods are used in clinical laboratories for identification of *Babesia, Trichomonas microti* in blood and from ticks, *Encephalitozoon* species microsporidia, and *Trichomonas vaginalis*. Multiplex PCR methods have been designed to simultaneously detect multiple parasites, for example, multiplex real-time PCR assays for detection of *Entamoeba histolytica, Giardia lamblia,* and *Cryptosporidium parvum* simultaneously or *E. histolytica, Giardia intestinalis,* and *Cryptosporidium* spp. simultaneously in stool samples. The multiplex PCR methods also include an internal control to determine the efficiency of the PCR and detect inhibition in the sample, which is likely in stool samples. The development of multiplex PCR assays to detect multiple parasites in stool samples is extremely useful. First, multiple parasites can cause diarrhea, and morphology is the only way to differentiate between causative agents. Second, patients can have multiple intestinal parasites at the same time, and laboratory detection of the presence of all parasites is important. Finally, multiple parasites are transmitted through the same source; thus, detection of all parasites and appropriate control measures will reduce large-scale outbreaks.

ANTIMICROBIAL AGENTS

Antimicrobial agents are of two types, those that inhibit microbial growth (*-static,* e.g., **bacteriostatic, fungistatic**) and those that kill organisms outright (*-cidal,* e.g., **bacteriocidal, fungicidal**). Antimicrobial agents for use in clinical applications are designed to be

TABLE 11.8 Mode of Action of Antimicrobial Agents

Mode of Action	Examples
Disrupts cell wall synthesis or integrity	**Beta-lactams** (penicillins and cephalosporins) Glycopeptides (vancomycin)
Disrupts cell membrane structure or function	Polymyxins (polymyxin B) Bacitracin
Inhibits protein synthesis	Aminoglycosides (gentamicin) Tetracyclines Macrolides (erythromycin) Lincosamides (clindamycin)
Inhibits nucleic acid synthesis or integrity	Quinolones (ciprofloxacin) Metronidazole
Inhibits metabolite synthesis	Sulfamethoxazole Trimethoprim

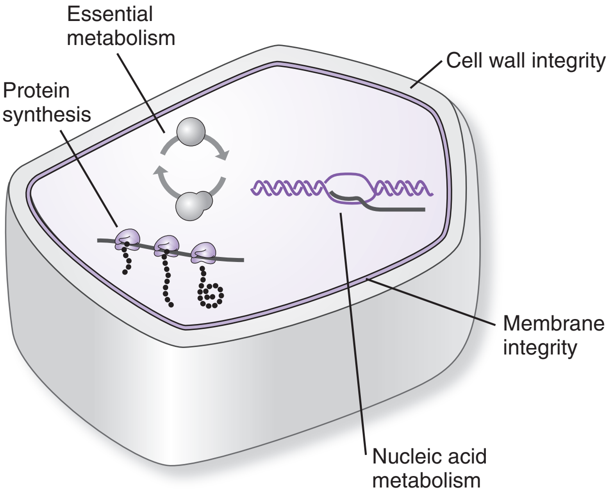


FIGURE 11.5 Sites of antimicrobial action. Depending on the type of organism, several structures can be affected by antimicrobial agents. All of these are essential for cell growth and survival.

selective for the target organism with minimal effect on mammalian cells. These agents are also intended to distribute well in the host and remain active for as long as possible (long half-life). Ideally, the agents should have -cidal (rather than -static) activity against a broad spectrum of microorganisms.

Another way to classify antimicrobial agents is by their mode of action (Table 11.8). The ultimate effect of these agents is to inhibit essential functions in the target organism (Fig. 11.5). A third way to group antimicrobial agents is by their chemical structure. For example, there are two major types of agents that inhibit cell wall synthesis, the β-lactams with substituted ring structures and the glycopeptides.

Resistance to Antimicrobial Agents

Microorganisms naturally develop defenses to antimicrobial agents. At the molecular level, resistance can arise from alerted target binding, active extrusion, or inaccessibility of the drug or microbial enzymes that inactivate the drug. Multidrug-resistant organisms may have one or more of these characteristics. A single-nucleotide variant in a drug target or transport protein can result in resistance. Resistant *Staphylococcus, Pseudomonas,* and *Klebsiella* spp. are becoming commonplace in health-care institutions. Long-term therapy with antibiotics such as vancomycin can lead to the development of resistant clones of organisms. These clones may persist in low numbers below the detection levels of routine laboratory sensitivity testing methods.

Advanced Concepts

The first antibiotics isolated were natural secretions from fungi and other organisms. Synthetic modifications of these natural agents were designed to increase the spectrum of activity (ability to kill more organisms) and to overcome resistance. For example, cephalosporins include such first-generation agents as cephalothin and cefazolin that are active against *Staphylococcus, Streptococcus,* and some Enterobacteriaceae. A second generation of cephalosporins—cefamandole, cefoxitin, and cefuroxime—is active against more Enterobacteriaceae and organisms resistant to β-lactam antibiotics. A third generation—cefotaxime, ceftriaxone, and ceftazidime—is active against *P. aeruginosa* as well as many Enterobacteriaceae and organisms resistant to β-lactam antibiotics. The fourth generation, cefepime, is active against an extended spectrum of organisms resistant to β-lactam antibiotics.

TABLE 11.9 Resistance Mechanisms

Mechanism	Example	Examples of Agents Affected
Destruction/ modification of agent	β-lactamases	β-lactams aminoglycosides
Elimination of agent	Multidrug efflux systems	β-lactams, fluoroquinolones, macrolides, chloramphenicol, trimethoprim
Altered cell wall structure	Thick cell walls that exclude agent Altered agent binding sites	Vancomycin β-lactams
Alternate metabolic pathways	Altered enzymes	Sulfonamides, trimethoprim

TABLE 11.10 Genes Conferring Resistance to Antimicrobial Agents in Particular Organisms

Organism	Antimicrobial Agent	Gene(s) Conferring Resistance
Staphylococcus aureus	Oxacillin	*mecA*
Streptococcus pneumoniae	Penicillin	*pbp1a* and *pbp1b*
Gram-negatives	β-lactams	*tem*, *shv*, *oxa*, *ctx-m*
Enterococcus	Vancomycin	*vanA*, *vanB*, *vanC*, *vanD*, *vanE*, *vanG*
Salmonella	Quinolones	*gyrA*, *gyrB*, *parC*, *parE*
Mycobacterium tuberculosis	Isoniazid	*katG*, *inhA*
	Rifampin	*rpoB*

There are several ways in which microorganisms develop resistance (Table 11.9). First, bacteria can produce enzymes that inactivate the agent. Examples of this resistance mechanism are seen in *S. aureus* and *N. gonorrhoeae* that produce β-lactamase, an enzyme that cleaves the β-lactam ring of the β-lactam antimicrobials such as the penicillins. Cleavage of the β-lactam ring destroys the activity of penicillin, rendering the organism resistant to its antimicrobial action. Second, organisms produce altered targets for the antimicrobial agent. Mutations in the gene encoding a penicillin-binding protein—for example, a change in the structure of the protein—render penicillin unable to bind to its target. Finally, bacteria exhibit changes in the transport of the antimicrobial agent either into or out of the cell. An example of this mechanism is seen in gram-negative bacteria that change their outer membrane proteins (porins) in order to decrease the influx of the antimicrobial agent. If the agent cannot get into the cell and bind to its target, then it is not effective in inhibiting or killing the bacterium.

All these resistance mechanisms involve a genetic change in the microorganism (Table 11.10). These genetic changes are most commonly brought about by mutation and selection processes. If a mutation results in a survival or growth advantage, cells with the mutation will eventually take the place of those without the mutation, which are less able to survive and procreate. This process is stimulated by antibiotic exposure, especially if the levels of antibiotics are less than optimal. For example, *S. aureus* developed resistance to antibiotics that target its penicillin-binding protein (PBP1) by replacing PBP1 with PBP2a encoded by the *mecA* gene. PBP2a found in methicillin-resistant *S. aureus* **(MRSA)** has a low binding affinity for methicillin.

Another genetic resistance mechanism is the acquisition of genetic factors from other resistant organisms through transformation with plasmids carrying resistance genes or transduction with viruses carrying resistance genes. Genetic factors can also be transferred from one bacterium to another by conjugation. Genetically directed resistance can pass between organisms of different species. For example, MRSA (vancomycin-sensitive *S. aureus*) can gain vancomycin resistance from vancomycin-resistant *Enterococcus faecalis*. Vancomycin and other glycopeptides act by preventing the cross-linking of the peptidoglycan, thereby inhibiting cell wall production. Several genes have been found in enterococci that encode altered binding proteins: *vanA, vanB, vanC, vanD, vanE*, and *vanG*. The expression of

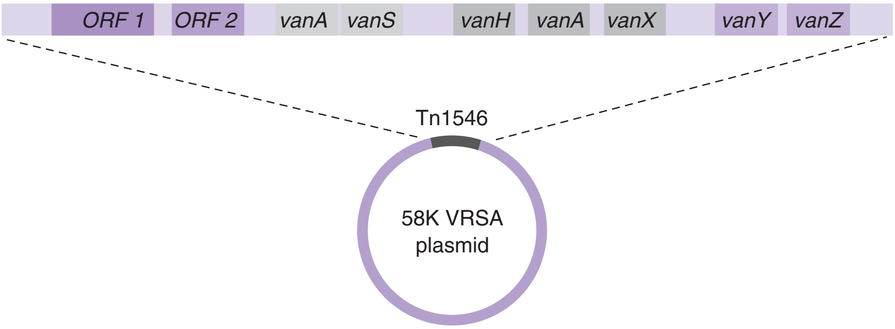


FIGURE 11.6 Vancomycin-resistant *S. aureus* (VRSA) plasmid carrying transposon Tn1546 with vancomycin resistance genes.

vanA and *vanB* is inducible and transferred from cell to cell by plasmids carrying vancomycin resistance genes on a **transposon**[55] (Fig. 11.6). The resulting vancomycin-resistant *S. aureus* (VRSA) uses lactic acid instead of alanine to build its cell wall. The VRSA cell wall, then, does not contain the target structure (D-ala-D-ala) for vancomycin. Development of resistance to particular drugs occurs with particular mutations: *rpoB* mutation is associated with rifampin resistance, and mutations in *katG, inhA, ahpC,* and *ndh* genes are associated with resistance to isoniazid.[56]

Molecular Detection of Resistance

Susceptibility to antimicrobial agents is determined by phenotypic or genotypic methods. The phenotypic development of resistance is detected by performing in vitro **susceptibility testing.** Testing for altered sensitivity to antimicrobial agents is of clinical significance especially when organisms persist in patients being treated with antimicrobial agents that are generally considered effective against the particular isolate or when large numbers of organisms are observed in normally sterile fluids, such as blood, cerebrospinal fluid, or urine. Phenotypic methods are generally used for aerobic bacteria, some mycobacteria, and yeast. For other organisms, such as viruses and filamentous fungi, phenotypic methods are not well standardized. Phenotypic methodologies include disk diffusion, broth dilution, and direct detection of resistance factors such as β-lactamase.

Susceptibility testing measures the **minimum inhibitory concentration (MIC)** of an antimicrobial agent, or the least amount of antimicrobial agent that inhibits the growth of an organism. Established guidelines define the MICs interpreted as indications of susceptibility or resistance for a given organism and antimicrobial agent pair. The determination of MICs to detect antimicrobial resistance is a phenotypic method. Although MIC methods are well established, and the results are generally reliable with regard to in vivo effectiveness of an agent for an organism, the results are sometimes difficult to interpret and the procedures are time-consuming, taking at least 48 hours after the specimen is collected.

Molecular methods that detect genes directly involved in the resistance of an organism to a particular agent offer a more straightforward prediction of antibiotic resistance. There are four reasons for using molecular-based methodologies. First, when an organism has an MIC at or near the breakpoint of resistance, detection of mutated genes contributing to resistance would be irrefutable evidence of the potential ineffectiveness of the agent. Second, genes involved in the resistance of organisms to antimicrobial agents can be detected directly in the clinical specimen closer to the time of collection and save the time required to isolate the organism and perform phenotypic MIC determinations on isolated colonies. With no requirement for culturing potentially dangerous microorganisms, there is less chance of hazardous exposure for the technologist as well. Third, monitoring the spread of a resistance gene in multiple isolates of the same organism is more useful in epidemiological investigations than following the trend in the MIC. Finally, molecular methods are considered the gold standard for validation of new phenotypic assays.

Beta-Lactam Antibiotic Resistance

The earliest commercialized antibiotic, penicillin, was first used therapeutically in the early 1940s (Fig. 11.7). Soon after that, resistance to penicillin through the

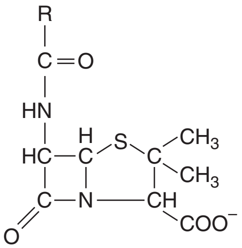


FIGURE 11.7 Structure of penicillin, showing the carbon-nitrogen (CCCN) beta-lactam ring.

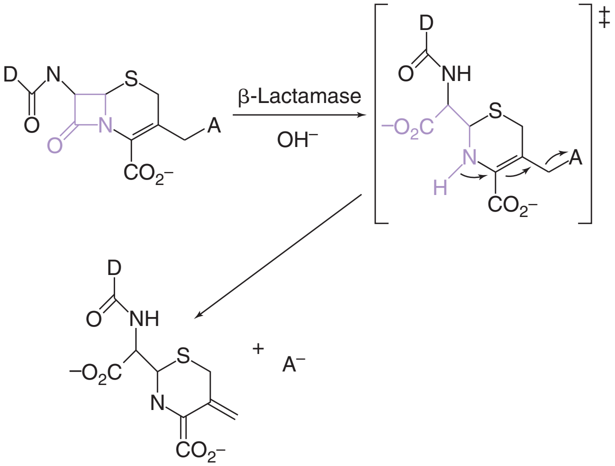


FIGURE 11.8 The reaction catalyzed by beta-lactamase results in cleavage of the beta-lactam ring through an intermediate product (bracketed).

production of β-lactamases (Fig. 11.8) was recorded. *Streptococcus pyogenes* is one of the very few organisms that are still predictably susceptible to penicillin. Modified β-lactam molecules including the cephalosporins and carbapenems were subsequently developed. Penicillin and the other β-lactam antimicrobials inhibit bacteria by interfering with an enzyme that is involved in the synthesis of the bacterial cell wall.

To counter the production of the bacterial β-lactamases (also known as penicillinases), penicillinase-resistant penicillins—for example, methicillin or oxacillin (Fig. 11.9)—were designed. Staphylococcal infections were treated successfully with methicillin/oxacillin until the emergence of resistance to these agents was first observed in 1965.[57] MRSA and methicillin-resistant coagulase-negative staphylococci became a major cause of infections acquired in the hospital as well as in the community. As described previously, expression of an altered penicillin-binding protein (PBP2' or PBP 2a encoded by the *mecA* gene) is the resistance mechanism. The antibiotics cannot bind to the altered target and therefore have no effect on the bacterial cells.

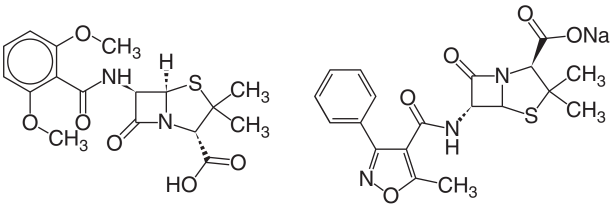


FIGURE 11.9 Penicillin substitutes include methicillin *(left)* and oxacillin *(right)*.

Rapid identification of MRSA isolates in clinical specimens by direct detection of *mecA* is critical for effective patient management and prevention of infections occurring in hospitals due to MRSA. PCR and other amplification assays have been developed for direct testing of clinical samples, and many assays have been tested for sensitivity and specificity.

The stereochemical structure of the carbapenems makes them more resistant to inactivation by most plasmid and chromosomal-mediated β-lactamases than the penicillins. After widespread use of these agents, however, resistance developed primarily through the production of carbapenem-hydrolyzing β-lactamases (carbapenemases).[58]

MALDI-TOF mass spectrometry peptide profiles can discriminate resistant strains of organisms such as MRSA and VRE. In a mass spectrometry β-lactamase assay, the antibiotic is mixed with the bacterial culture. After incubation, the bacteria are removed by centrifugation, and the supernatant is analyzed for the hydrolyzed form of β-lactam. This method can detect resistance to penicillin and a variety of β-lactam antibiotics, including ampicillin, piperacillin, ceftazidime, imipenem, and others.[59] Modifications of the method have been aimed at increased speed and accuracy. Carbapenemase production from anaerobic bacteria in less than 3 hours has been reported.[60,61]

Glycopeptide Antibiotic Resistance

Glycopeptide antibiotics were originally isolated from plant and soil bacteria. Their structures contain either

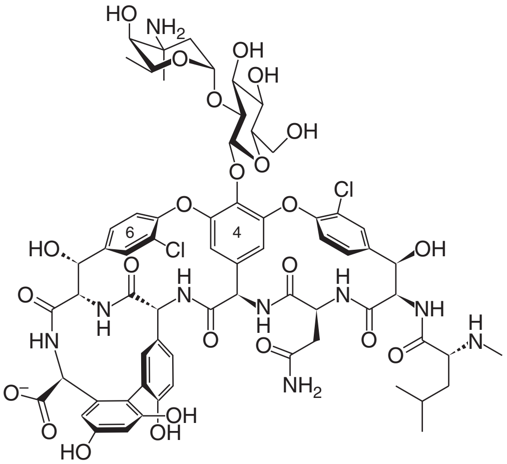


FIGURE 11.10 Structure of vancomycin.

a glycosylated cyclic or polycyclic peptide. These antibiotics inhibit peptidoglycan synthesis in the cell wall of susceptible organisms (principally gram-positive cocci). Glycopeptide antibiotics include vancomycin (Fig. 11.10), teicoplanin, and ramoplanin; oritavancin, dalbavancin, and telavancin are semisynthetic glycopeptide antibiotics. *Enterococcus* was the first organism in which glycopeptide resistance was reported.[62] Since then, vancomycin resistance has been observed in other organisms.

The mechanism of vancomycin resistance can be acquired or intrinsic to the organism. The antibiotic binds and weakens cell wall structures of susceptible bacteria, ultimately causing cell lysis. Either the presence of the antibiotic itself or the weakening of the cell wall activates a set of *van* genes whose products modify the cell wall structure such that the antibiotic no longer binds. The *van* genes are carried on a cassette localized to a plasmid or on the bacterial chromosome. Resistance arises from enzymatic modifications of the aminoglycosides by acetyl-, adeno-, and phospho-transferases. A universal database for the detection of the modified drugs by MALDI-TOF MS has been established. To date, these methods are not yet in routine laboratory use because they have not achieved the ease and reproducibility of routine diagnostic systems.[63]

PCR was used to detect the *vanA, vanB, vanC1,* and *vanC2* resistance genes in fecal samples as a way to screen for vancomycin-resistant enterococci (VRE). The specificity of PCR and multiplex PCR was high compared with the isolation of VRE in culture. Quantitative PCR has also been used to detect VRE in fecal surveillance specimens. Because PCR is faster than traditional culture methods and has comparable sensitivity and specificity to culture, it is an attractive method for screening large numbers of samples for a particular target.

Antimicrobial Resistance in M. tuberculosis

The use of molecular methods greatly improved antimicrobial resistance detection in *M. tuberculosis* because traditional methods of determining antimicrobial susceptibility took days, if not weeks, for this organism. The longer a patient with tuberculosis (TB) is inadequately treated, the more likely the development of resistance and possible spread of the resistant organisms. Evolution of drug resistance of *M. tuberculosis* in a noncompliant patient for over 12 years was reported where subpopulations of the organism emerged due to the acquisition and accumulation of gene mutations that rendered the organism resistant to isoniazid, rifampin, and streptomycin.[64] Many nucleic acid amplification protocols have been developed to directly detect mutations in the genes associated with resistance to isoniazid and rifampin. In general, these assays have demonstrated excellent sensitivity and specificity and provide rapid determination of drug susceptibility either directly from sputum or from cultures.

Thus, the advantages of using nucleic acid amplification assays for the determination of drug resistance include the rapid and specific detection of mutations in genes associated with resistance to particular antimicrobial agents that provides irrefutable evidence of resistance in a short period.

MOLECULAR EPIDEMIOLOGY

An **epidemic** is a disease or condition that affects many unrelated individuals at the same time; a rapidly spreading outbreak of an infectious disease is an epidemic. A **pandemic** is a disease that sweeps across wide geographical areas. Epidemiology includes collection and analysis of environmental, microbiological, and clinical data. In microbiology, studies are performed to follow the spread of pathogenic organisms within a hospital (**nosocomial** infections), from the actions of a physician (**iatrogenic** infections), and in a community. Molecular epidemiology is the study of causative genetic and environmental factors at the molecular level to ascertain the origin, distribution, and best strategies for prevention of disease. In infectious disease, these efforts are facilitated by the ability to determine the genetic similarities and differences among microbiological isolates.

In the laboratory, molecular methods are very useful for identifying and typing infectious agents. In a single patient, this is informative for therapeutic efficacy; in groups of patients, it provides information for infection control. Typing systems are based on the premise that clonally related isolates share molecular characteristics distinct from unrelated isolates. Molecular technology provides analytical alternatives from the chromosomal to the nucleotide sequence level. These genotypic methods, in addition to established phenotypic methods, enhance the capability to identify and type microorganisms. Whereas phenotypic methods are based on a range of biological characteristics, such as antigenic type or growth requirements, genotypic procedures target genomic or plasmid DNA (Table 11.11). Genome scanning methods, such as restriction enzyme analysis followed by PFGE, are used to find genetic similarities and differences, as are amplification and sequencing methods. Mass spectrometry databases continue to accumulate for detailed typing of bacteria and fungi by peptide profiles. The ability to discern differences with increasing detail enhances the capability to type organisms regardless of their complexity. All methods, however, have benefits and limitations with regard to instrumentation, methodology, and interpretation.

TABLE 11.11 Epidemiological Typing Methods

Class	Methods
Phenotypic	Biotyping, growth on selective media Antimicrobial susceptibility Serotyping, immunoblotting Bacteriophage typing Protein, enzyme typing by electrophoresis MALDI-TOF mass spectrometry
Genotypic	Plasmid analysis Restriction endonuclease mapping Pulsed-field gel electrophoresis Ribotyping Arbitrarily primed PCR, RAPD PCR Melt-curve analysis REP-PCR, ERIC PCR, ITS, *spa* typing Mass parallel sequencing

Molecular methods are based on DNA and amino acid sequences. Nucleotide and amino acid sequences range from highly conserved across species and genera to unique to each organism. Some of these are strain- or species-specific sequences but are still used for epidemiological testing because these methods are highly reproducible and, depending on the targets, can discriminate between even closely related organisms. Most (but not all) molecular methods offer definitive results in the form of DNA sequences, peptide profiles, or gel band and peak patterns that can be interpreted objectively, which is less difficult than phenotypic determinations that often require experienced judgment.[65] With commercial systems, the performance has become relatively simple for some molecular epidemiology tests, whereas others require a higher level of laboratory expertise.

Molecular Strain Typing Methods for Epidemiological Studies

In community or clinical settings, the same organism might be isolated multiple times, whether in the same patient or from different patients. The physician has to determine whether collected isolates were independently acquired, that is, came from different sources, or if they came from the same source. With this knowledge, the actions are taken to control the transmission of the organism, especially if it is being transmitted from a common source and that source has been identified. Most of the time, these analyses are performed on organisms that have been transmitted nosocomially, but sometimes procedures to determine relatedness are performed on isolates from community outbreak situations.[66]

Many laboratory methods can be used to determine the relatedness of multiple isolates, both phenotypic (e.g., by MALDI-TOF, serology, and antimicrobial susceptibility patterns) and genotypic (e.g., PFGE and ribotyping). The phenotypic methods suffer from a lack of reproducibility and their ability to discriminate between isolates. Even after recent advances in mass spectrometry, genotypic methods are used almost exclusively to type bacterial strains to determine the relatedness of multiple isolates.

Plasmid Analysis

Plasmid analysis involves isolation and restriction mapping of bacterial plasmids. The same bacterial strain can have different plasmids carrying different phenotypes or resistance patterns. For this analysis, plasmid DNA is isolated from the specimen or culture and then digested with restriction enzymes. Plasmids are distinguished by gel electrophoresis patterns of fragments generated when cut with the appropriate enzymes. Restriction analysis can also be performed on chromosomal DNA of organisms with small genomes. For organisms with larger genomes, whole genome analysis with restriction enzymes that cut frequently enough to identify plasmids will yield complex patterns that are more difficult to interpret.

Pulsed-Field Gel Electrophoresis

Most molecular epidemiological tests are performed using PFGE, which can identify organisms with larger genomes or multiple chromosomes. For PFGE analysis, the DNA is digested with restriction enzymes that cut infrequently within the genomic sequences. The resulting large fragments (hundreds of thousands of base pairs) are resolved by PFGE. Banding patterns will differ depending on the chromosomal DNA sequence of the organisms (Fig. 11.11). Tenover and colleagues devised a system to interpret the banding pattern of a test organism compared to that of the strain of an identified or reference organism.[67] The interpretation of PFGE results follows the "rule of three" (Table 11.12), a method that has been used for typing numerous species, including strains of *Pseudomonas aeruginosa, Mycobacterium avium, Escherichia coli, N. gonorrhoeae*, VRE, and MRSA. Intralaboratory and interlaboratory computerized databases of band patterns can be stored for reference. A national PFGE database is stored at the CDC (www.cdc.gov/pulsenet). One disadvantage of using PFGE to type strains is the time involved to perform the assay; it can take 2 to 3 days to complete one analysis.

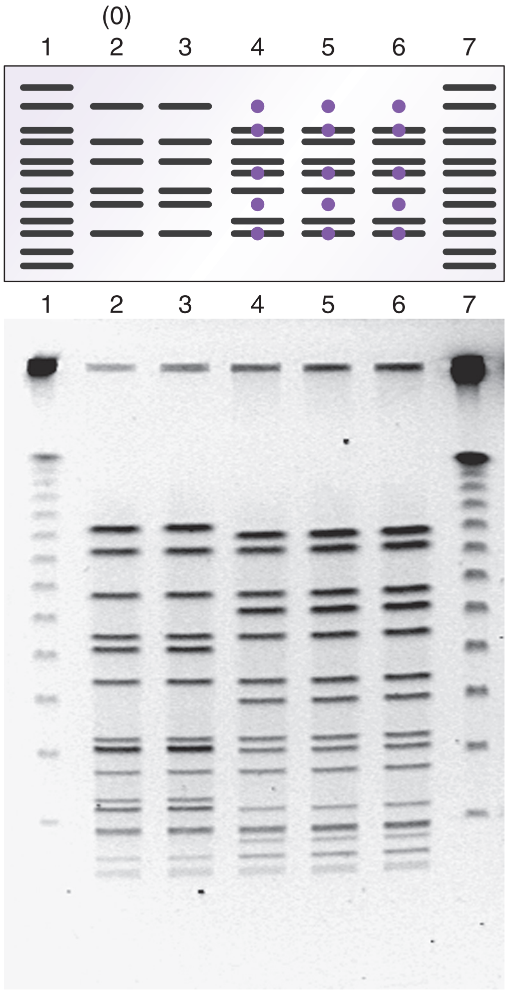


FIGURE 11.11 PFGE of coagulase-negative *Staphylococcus* showing the outbreak (O) and four test strains *(top panel)*. The gel image is shown in the bottom panel. The fragment shifts are marked in the top panel. Strains in lanes 4, 5, and 6 are the same but different from the outbreak strain (lane 2). The strain in lane 3 is the same as the outbreak strain. The strains in lanes 4, 5, and 6 have at least five genetic differences and are not related to the outbreak. Lanes 1 and 7, molecular-weight markers. *(Courtesy of Mary Hayden, MD, Rush Medical Laboratories, Rush University Medical Center, Chicago, IL.)*

Restriction Fragment Length Polymorphism Analysis

RFLP analysis by Southern blot is the same technique first used to identify and investigate human genes. This

TABLE 11.12 Criteria for PFGE Pattern Interpretation

Category	Genetic Differences*	Fragment Differences*	Epidemiological Interpretation
Indistinguishable	0	0	Test isolate is the same strain as the outbreak strain
Closely related	1	2–3	Test isolate is closely related to the outbreak strain
Possibly related	2	4–6	Test isolate is possibly related to the outbreak strain
Different	≥3	≥6	Test isolate is unrelated to the outbreak

*Compared with the outbreak strain.

method involves cutting DNA with restriction enzymes, resolving the resulting fragments by gel electrophoresis, and then transferring the separated fragments to a membrane for probing with a specific probe. Gene-specific probes are used to identify or subtype microorganisms such as *P. aeruginosa* in patients with cystic fibrosis and nosocomial *L. pneumophila* infections.

For strain typing of *M. tuberculosis* by RFLP, the probe is complementary to an insertion element, IS*6110*, and will bind to restriction fragments on the membrane that contains it, resulting in a series of bands that are easily analyzed and compared. Insertion elements (insertion sequences) are segments of DNA that move independently throughout the genome and insert themselves in multiple locations. Strains can be RFLP-typed based on how many insertions are present and where they are located. Isolates from the same strain will have the same number and location of elements.

The gene targets selected for RFLP depend on the organism under investigation and which genes will be most informative. Ribosomal RNA genes are highly informative over a range of microorganisms, which is the basis for a modification of the RFLP procedure, which is a form of **ribotyping.** For this method, probes target the 16S and 23S rRNA genes. RFLP and ribotyping have been applied in industrial as well as clinical microbiology.

RFLP can be performed more rapidly using PCR amplification with gene-specific primers **(locus-specific RFLP [PCR-RFLP]).** This method requires the amplification of specific regions by PCR. The amplicons are then cut with restriction enzymes, yielding bands of informative size. An advantage of this procedure, in addition to its speed, is the simple banding pattern, which is much easier to interpret. Although the method is limited by the sequences that can be amplified and differentiated through restriction enzyme digestion, proper gene selection provides a highly reproducible and discriminatory test. In one study, an 820-bp amplified fragment of the *ureC* gene from *Helicobacter pylori* digested with *Sau*3A and *Hha*I yielded 14 different *Sau*3A patterns and 15 different *Hha*I patterns. These patterns were informative as to the antibiotic sensitivity of the various types to clarithromycin therapy.

Arbitrarily Primed PCR

Arbitrarily primed PCR, or random amplified polymorphic DNA (RAPD) assay, is a modified PCR using short (e.g., 10-base-long) oligonucleotides of random sequences to prime DNA amplification all over the genome. The gel pattern of amplicons produced is characteristic of a given organism. If two organisms have the same pattern, they are considered the same type; if the patterns differ, they are different types. The RAPD assay is relatively fast and inexpensive; however, producing consistent results may be technically demanding. Accurate interpretation of RAPD raw data requires that the procedure conditions be followed strictly so that pattern differences (not necessarily patterns) are reproducible (Fig. 11.12).

Amplified Fragment Length Polymorphism (AFLP) Assay

AFLP is a name, rather than an acronym, chosen by the inventors of this assay due to its resemblance to RFLP. The AFLP assay is based on the amplification of DNA fragments generated by cutting the test genome with

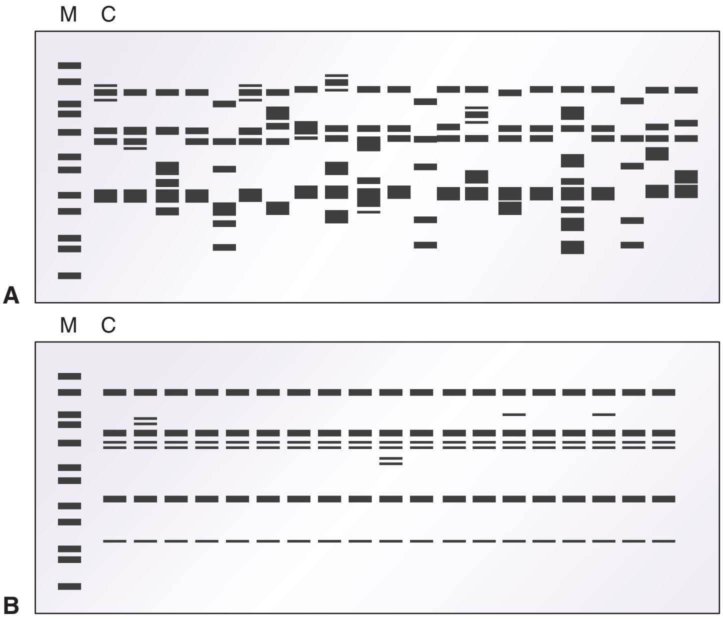


FIGURE 11.12 RAPD gel results. An unacceptable gel pattern is represented in panel A. The bands are smeared, and variable producing patterns are too complex for positive identification of unrelated strains. The gel pattern represented in panel B is acceptable. Strain differences can be clearly identified by variations from the known strain (C). Molecular-weight markers are shown in lane M.

restriction enzymes. DNA isolated from the test strain is digested with *Hind*III or other restriction enzyme (Fig. 11.13). Short DNA fragments or adaptors are ligated to the ends of the cut DNA. The adaptor-ligated fragments are then amplified in two steps with primers complementary to the adaptor sequences. Nucleotides located at the 3′ end of the primers select for specific sequences in the restriction fragments. The amplicons are then resolved by gel or capillary electrophoresis (fluorescently labeled primers are used for capillary electrophoresis). The pattern will be characteristic of the strain or type of organism. This assay can be performed with one or two enzymes (e.g., *Eco*R1/*Mse*I or *Bam*H1/*Pst*I).

AFLP may detect more polymorphisms than RAPD analysis and is faster than PFGE. The procedure is more technically demanding, however, than a similar method, **REP-PCR** (discussed in the next section). Gel patterns may also be complex (Fig. 11.14). Both high and low reproducibility for the method have been reported.

Interspersed Repetitive Elements

Copies of conserved sequences are found throughout the genomes of most organisms. These sequences may have arisen from viral integration or movement of transposable elements (stretches of DNA that move from one location to another in a non-Mendelian fashion; also called jumping genes). The genomic locations of these structures are related to species type and can be used to distinguish between bacterial isolates.

Enterobacterial repetitive intergenic consensus (ERIC) sequences are 126-bp-long genomic sequences found in some bacterial species that are highly conserved, even though they are not in coding regions. These sequences are located between genes in operons or upstream or downstream of single open reading frames. ERIC sequences are flanked by inverted repeats that could form stem-loop or cruciform structures in DNA and are found only in gram-negative organisms, such as *Bartonella, Shigella, Pseudomonas, Salmonella, Enterobacter,* and others.

A related type of repetitive element, the repetitive extragenic palindromic (REP) sequence, is similar to the ERIC sequence in that it occurs in noncoding regions and contains an inverted repeat. REP sequences differ from ERIC sequences in size (REP sequences are only 38 bp long), in being more numerous in the genome, and in being present in multiple copies at a single location (Fig. 11.15). PCR primed from these elements yields a series of products that can be resolved by gel, capillary

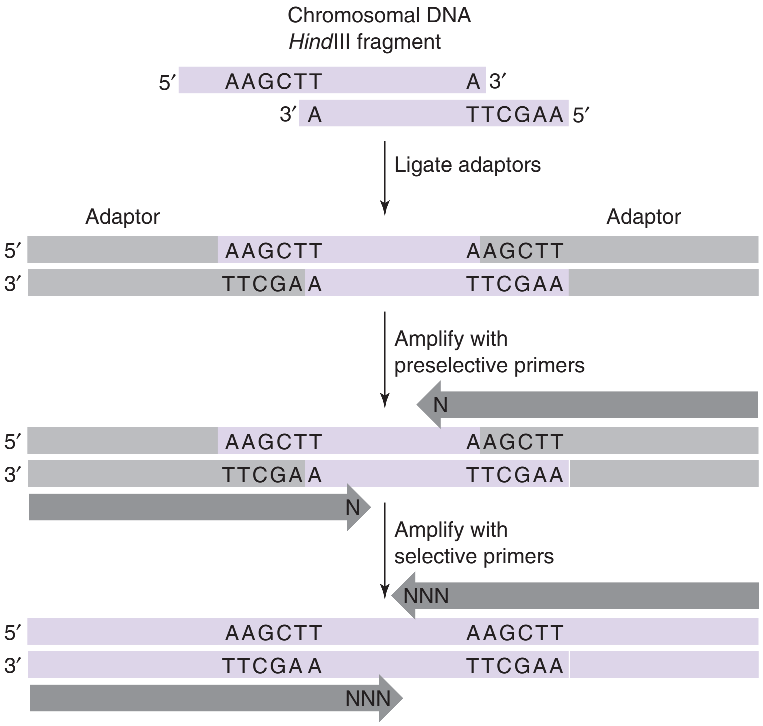


FIGURE 11.13 AFLP analysis begins with restriction digestion of chromosomal DNA. The resulting fragments *(top)* are ligated with adaptors compatible with the restriction enzyme ends and complementary to primers used to amplify them. The first amplification is performed with preselective primers that end in a 3′ base (N) selected by the user. Selective primers with three added 3′ bases are used for a second round of PCR. This selection results in a characteristic pattern; only a fraction of the original fragments will be represented in the gel pattern.

electrophoresis, or microfluidics into characteristic patterns (Fig. 11.16). These elements have been used for typing of clinically important organisms, such as *Clostridium difficile* and fungal pathogens.

Another repetitive element, **BOX,** was discovered in *S. pneumoniae.* BOX elements consist of different combinations of subunits, boxA, boxB, and boxC, which are 59, 45, and 50 bp long, respectively. Although these elements are not related to ERIC and REP sequences, they do form stem-loop structures, as do ERIC and REP. About 25 of these elements are present in the *S. pneumoniae* genome, where they may be involved in regulation of gene expression.

Internal Transcribed Spacer Elements

The ribosomal RNA genes comprise the most conserved region in the genome. These genes are arranged as an operon, including a small subunit, 18S rRNA, 5.8S rRNA, and a large subunit, 28S rRNA. The ITS 1 and 2 elements (ITS1 and ITS2) are found in regions separating the 18S and the 28S rRNA genes. ITS1 is located between the 18S and the 5.8S gene, and ITS2 is located between the 5.8S and the 28S rRNA genes. Two additional elements, intergenic spacer (IGS) regions, IGSI and IGSII, are located between the rDNA repeat units (Fig. 11.17). Used for the identification and typing of yeast and molds, ITS sequences are conserved within species but polymorphic between species. The ITS sequences are amplified using primers directed to the unique 17S and 26S gene sequences. The resulting amplicons are analyzed by sequencing, single-strand conformation polymorphism, density-gradient gel electrophoresis, restriction enzyme analysis, or sequence-specific PCR.

spa *Typing*

MRSA contains a VNTR element in the 3′ coding region of the protein A gene *(spa)*. The element consists of repeat units of 21 or 24 bp in length. Repeat units also vary by sequence of at least one position. The *spa* element can have 2 to 16 repeat units. Analysis of these elements by PFGE or sequencing and comparison to known isolates **(*spa* typing)** is used to identify MRSA. Almost 400 repeat profiles or *spa* types have been defined.

A similar sequence structure in the coagulase gene *(coa)* has also been used for *S. aureus* typing (***coa* typing**). The *coa* VNTR units are 81 bp in length. PCR amplification and sequence analysis are used to analyze *coa* types. The discriminatory power of this method is increased by analysis of other repeated sequences elsewhere in the MRSA genome. The combination of *spa* and VNTR typing has a discriminatory power equal to that of PFGE, with a more rapid turnaround time.

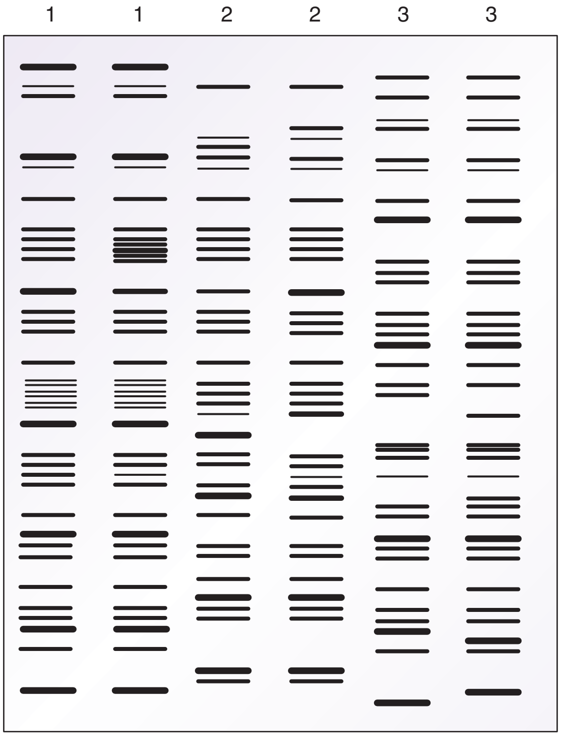


FIGURE 11.14 Banding patterns generated by fluorescent AFLP analysis. Note that duplicate specimens (1, 2, and 3) do not produce the exact pattern because of band shifts and different band intensities.

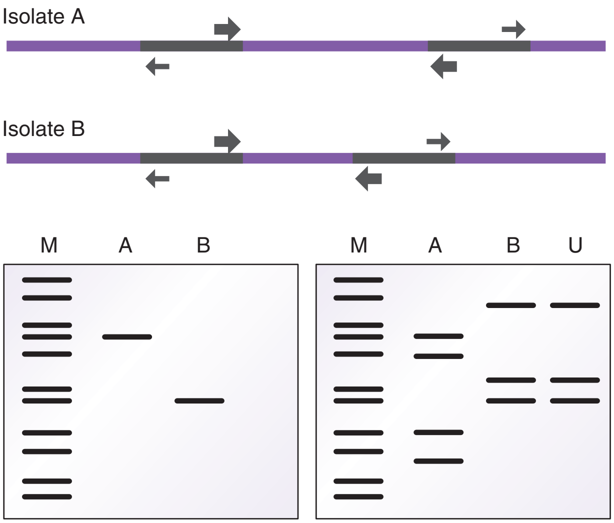


FIGURE 11.16 Species identification by REP or ERIC primed amplification. REP and ERIC sequences are present in different chromosomal locations in bacterial subtypes A and B. PCR, which extends outward-oriented primers hybridized to the repeated sequences, generates amplicons of different sizes based on the placement of the element, as in the gel depicted in the bottom left panel where only one amplicon is shown. Multiple REPs and ERICs throughout the bacterial chromosome generate multiple amplicons with characteristic gel patterns *(bottom right)*. An unknown isolate (U) is identified as the same strain as isolate B. *M,* molecular-weight markers.

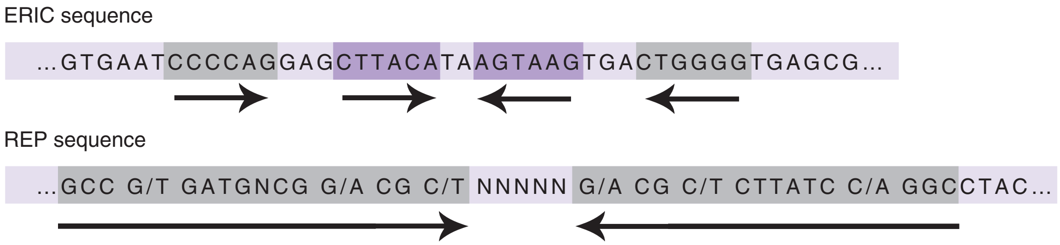

ERIC sequence
...GTGAATCCCCAGGAGCTTACATAAGTAAGTGACTGGGGTGAGCG...
REP sequence
...GCC G/T GATGNCG G/A CG C/T NNNNN G/A CG C/T CTTATC C/A GGCCTAC...

FIGURE 11.15 The central inverted repeat of a 126-bp ERIC sequence *(top)* and a consensus REP sequence *(bottom)*. ERIC sequences contain multiple inverted repeats (arrows) that can generate a secondary structure consisting of several stems and loops. REP sequences contain a conserved inverted repeat that forms a stem with a loop that includes 5 bp of variable sequence (N).

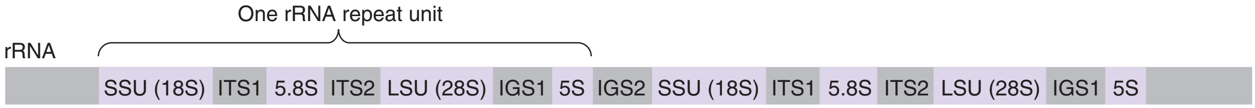


FIGURE 11.17 Ribosomal RNA genes are arranged in multiple tandem units that include the major rRNA transcript (18S to 28S) and the 5S gene. ITSs are located within the major transcript template area and IGSs in the region between the repeat units surrounding the 5S rRNA gene.

TABLE 11.13 Example of Housekeeping Genes Sequenced in an MLST Test

Gene	Gene Product	Number of Alleles
ArcC	Carbamate kinase	52
AroE	Shikimate dehydrogenase	88
GlpF	Glycerol kinase	55
Gmk	Guanylate kinase	51
Pta	Phosphate acetyltransferase	57
Tpi	Triosephosphate isomerase	74
YqiL	Acetyl coenzyme A acetyltransferase	66

Multilocus Sequence Typing

Multilocus sequence typing (MLST) characterizes bacterial isolates by using sequences of internal fragments of housekeeping genes. Six or seven genes are sequenced over 450 to 500 bases, and the sequences are assigned as alleles. Examples of housekeeping genes used in *S. aureus* MLST are shown in Table 11.13. Distinct alleles are defined as single-base-pair differences or multiple changes resulting from recombination or other genetic events. An isolate type or the allelic profile is the collection of all seven alleles, also called the sequence type. Bacteria have enough variation so that there are multiple alleles of each housekeeping gene, making up billions of allelic profiles. Because the data generated by MLST is text sequence, the results can be compared with those in large databases, for example, http://pubmlst.org.

Mass Spectrometry

Peptides (2 to 20 kd) of the highly abundant ribosomal proteins are assessed to identify microbes by MALDI-TOF peak patterns or peptide mass profiles.[68] The profiles are analyzed by searching for matching profiles in extensive open-source and intralaboratory databases. MALDI methods are faster than some genetic methods and can successfully detect organisms from mixtures. Use of MALDI results for stain comparisons is limited, however. Peptide profiles are influenced by growth conditions which affect microbial physiology and protein expression. Although identification by mass spectrometry requires culture of microorganisms, culture conditions and time of culture do not seem to alter profiles.[69,70]

Speciation of organisms that cause environmental hazards, such as food- and water-borne diseases, is important not only for investigating but also for preventing outbreaks. Only some strains of organisms are implicated in outbreaks. Peptide maps of organisms, such as *Aeromonas,* from which 7 of 17 species are related to water outbreaks, allow accurate classification of species for environmental monitoring.[71] MALDI has also been applied to the identification of food pathogens such as lactic acid bacteria from spoiled food, toxin-producing bacteria responsible for food poisoning, and beneficial bacteria in probiotics and yogurt.

Whole-cell MALDI-TOF mass spectrometry can be used to identify and compare organisms from highly contaminated ecosystems, such as soil and sewage sludge.[72] In this application, intact cells are extracted with trifluoroacetic acid and acetonitrile and permeabilized with glass beads before ionization. This method is also applied to intact conidia and spores for identification of fungal organisms.

Comparison of Typing Methods

Genotypic methods used for strain typing are evaluated and compared based on five criteria. First, the target organism must be typable by the method used to detect it **(typing capacity).** A test that detects a genotypic or

TABLE 11.14 Performance Comparison of Representative Molecular Epidemiology Methods*

Method	Typing Capacity	Discriminatory Power	Reproducibility†	Ease of Use	Ease of Interpretation‡
Plasmid analysis	Good	Good	Good	High	Good
PFGE	High	High	High	Moderate	Good–Moderate
Chromosomal RFLP	High	Good	Good	High	Moderate–Poor
Ribotyping	High	High	High	Good	High
PCR-RFLP	Good	Moderate	Good	High	High
RAPD	High	High	Poor	High	Good–High
AFLP	High	High	Good	Moderate	High
Repetitive elements	Good	Good	High	High	High
Sequencing	High	High	High	Moderate	Good–High
Mass spectrometry	High	Good-High	Good	Good	High

*From Olive and Bean.[76]

†Intralaboratory

‡Interpretation is influenced by the quality of the data.

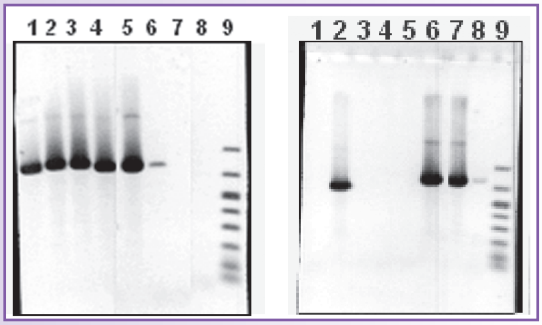
phenotypic characteristic that is not expressed in all members of a species will not accurately detect the target organisms at all times. A molecular assay must target a reasonably polymorphic DNA sequence, alleles of which are unambiguously associated with a given strain. Second, the method must be **reproducible.** A reproducible method yields the same result on repeated testing of the same organism or strain. Variations in cell characteristics, such as antigens or receptor expression, decrease reproducibility (precision). Third, the method must clearly distinguish between unrelated strains **(discriminatory power).** Reproducibility and discriminatory power are important for the establishment of databases that can be used by independent laboratories. Fourth, ease of interpretation of results is important. Unclear or complex results will lower reproducibility and discriminatory power. Finally, ease of test performance is important in order to minimize the chance of error or ambiguous results. A rating of representative methods is shown in Table 11.14.

The most desirable typing method is the one that will type all strains and have excellent reproducibility, discriminatory power, and ease of performance and interpretation. Unfortunately, no such method fits this profile, so the laboratory professionals performing these types of analyses may have to sacrifice ease of performance, for example, in order to get excellent discriminatory power when they are choosing which molecular typing method to use.

In conclusion, molecular-based methods are available for the detection, identification, and characterization of a number of microorganisms. For some, assays are used almost exclusively, such as for *N. gonorrhoeae, C. trachomatis,* and *B. pertussis*. For others, molecular-based tests are used to provide rapid results and supplement traditional testing, such as for *M. tuberculosis*. Finally, molecular-based methods are undergoing continuous development and experiencing increasing use compared to traditional culture and phenotypic methods.

Case Study 11.1

During a holiday weekend at a luxury hotel, guests began to complain of stomach flu with nausea and vomiting. In all, more than 100 of the 200 guests who had dined at the hotel the previous evening described the same symptoms. Eight people had symptoms severe enough to warrant hospitalization. Most, however, recovered within 24 to 48 hours of the onset of symptoms. Health officials were notified. Interviews and epidemiological analyses pointed to a Norwalk-like virus infection, or norovirus, probably foodborne. Stool specimens (1 to 5 mL) from hospitalized patients and samples of the suspected food sources (500 mg) were sent for laboratory analysis by RT-PCR. RNA extracted with 1,1,2 trichloro-1,2,2, trifluoroethane was mixed with a guanidium thiocyanate buffer and isolated by organic (phenol-chloroform) procedures. cDNA was synthesized using primers specific to the viral RNA polymerase gene, and strain-specific PCR primers were used to amplify the viral gene. The amplicons, resolved by agarose gel electrophoresis, are shown in the accompanying figure.



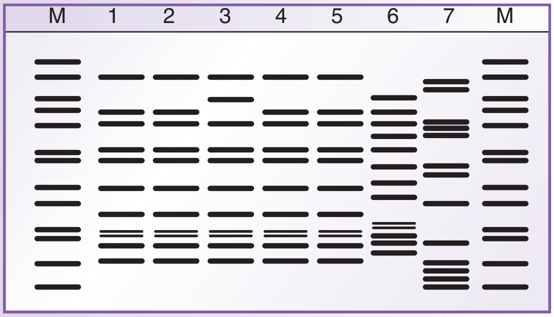
RT-PCR products were resolved on separate gels. Amplicons from four affected individuals (lanes 1 to 4, *left*). Lane 5, positive control; lane 6, sensitivity control; lane 7, negative control; lane 8, reagent blank. Specimens from suspected food sources (lanes 1 to 4, *right*). Lane 5, salad lettuce from the distributor; lanes 6 to 8, specimens from three hotel employees working the day of the outbreak; lane 9, molecular-weight marker.

QUESTIONS:

1. *Do the four patients have norovirus?*
2. *What is the source of the organism?*
3. *When did the source get contaminated? Did the source come into the hotel infected, or was it infected inside the hotel?*

Case Study 11.2

Five students from different local community high schools suffered recurrent skin infections with chronic wounds. Nasal swabs and skin specimens from the students were screened for MRSA by inoculation onto Mueller–Hinton agar supplemented with NaCl (0.68 mol/L) containing 6 µg/mL of oxacillin. The cultured organisms exhibited an MIC of more than 32 µg/mL vancomycin and a zone of inhibition of less than 14 mm in diameter. Isolates were sent to the CDC and referred for molecular testing. DNA isolated from the five cases and a control strain were embedded in agarose plugs and digested with SmaI for PFGE analysis. A depiction of the gel pattern is shown in the accompanying figure.



PFGE patterns of isolated strains. M, molecular-weight markers. Lanes 1 to 5, isolates from the community infections; lanes 6 and 7, unrelated hospital isolates.

Further PCR testing was performed on the isolates for virulence factors, particularly for *mecA* gene sequencing and detection of the Panton–Valentine leukocidin (PVL) genes, *lukS-PV* and *lukF-PV*. The results from these tests revealed that all five isolates contained the PVL genes and the type IV *mecA* element.

QUESTIONS:

1. *Are all or some of the five isolates the same or different? Which isolates are the same, and which are different? What is the evidence to support your answer?*
2. *Was there a single source for these organisms or multiple sources?*

Case Study 11.3

A 39-year-old HIV-positive male has been monitored closely since his diagnosis as HIV-positive 5 years ago. The man was on HAART and compliant. He was relatively healthy and had not even had a cold in the last 4 years. The man had HIV viral loads as determined by Amplicor RT-PCR that were consistently close to 10,000 copies/mL, never varying more than 0.2 $\log_{10}$ unit, until the last 6 months, when his viral loads were trending up to 25,500, then 48,900, with his most recent result being 55,000 copies/mL. Genotyping performed on virus isolated from the patient revealed a mutation in the reverse transcriptase gene of M41L that is associated with resistance to zidovudine (AZT).

QUESTIONS:

1. *What is the significance of the viral load results over the last 6 months?*
2. *What is the implication of the genotyping result for the patient's therapy?*
3. *How should this patient be monitored in the future?*

STUDY QUESTIONS

1. Which of the following genes would be analyzed to determine whether an isolate of *Staphylococcus aureus* is resistant to oxacillin?

a. *mecA*
b. *gyrA*
c. *inhA*
d. *vanA*

2. Which of the following is a genotypic method used to compare two isolates in an epidemiological investigation?

a. Biotyping
b. Serotyping
c. Ribotyping
d. Bacteriophage typing

3. For which of the following organisms must caution be exercised when evaluating positive PCR results because the organism can be found as normal flora in some patient populations?

a. *Neisseria gonorrhoeae*
b. HIV
c. *Chlamydophila pneumoniae*
d. *Streptococcus pneumoniae*

4. Which of the following controls are critical for ensuring that amplification is occurring in a patient sample and that the lack of PCR product is not due to the presence of polymerase inhibitors?

a. Reagent blank
b. Sensitivity control
c. Negative control
d. Amplification control

5. A PCR assay was performed to detect *Bordetella pertussis* on sputum obtained from a 14-year-old girl who has had a chronic cough. The results revealed two bands, one consistent with the internal control and the other consistent with the size expected for amplification of the *B. pertussis* target. How should these results be interpreted?

a. These are false-positive results for *B. pertussis.*
b. The girl has clinically significant *B. pertussis* infection.
c. *B. pertussis* detection is more likely due to colonization.
d. The results are invalid because two bands were present.

6. Which of the following is an advantage of molecular-based testing?

a. Results stay positive long after successful treatment.
b. Results are available within hours.
c. Only viable cells yield positive results.
d. Several milliliters of specimen must be submitted for analysis.

7. Which molecular-based typing method has high typing capacity, reproducibility and discriminatory power, moderate ease of performance, and good-to-moderate ease of interpretation?
   - a. Repetitive elements
   - b. PFGE
   - c. Plasmid analysis
   - d. PCR-RFLP

8. A patient has antibodies against HCV and a viral load of 100,000 copies/mL. What is the next test that should be performed on this patient's isolate?
   - a. Ribotyping
   - b. PCR-RFLP
   - c. Hybrid capture

9. A positive result for HPV type 16 indicates
   - a. high risk for antibiotic resistance.
   - b. low risk for cervical cancer.
   - c. high risk for cervical cancer.

10. Which of the following is used to type molds?
    - a. Sequence-specific PCR
    - b. Microarray
    - c. ITS sequencing
    - d. Flow cytometry

References

1. Boulund F, Karlsson R, Gonzales-Siles L, Johnning A, Karami N, Al-Bayati O, Åhrén C, Moore ERB, Kristiansson E. Typing and characterization of bacteria using bottom-up tandem mass spectrometry proteomics. *Molecular and Cellular Proteomics* 2017;16:1052–1063.
2. Akand E, Downard KM. Mutational analysis employing a phylogenetic mass tree approach in a study of the evolution of the influenza virus. *Molecular Phylogenetic Evolution* 2017;112:209–217.
3. Sefers S, Pei Z, Tang Y-W. False positives and false negatives encountered in diagnostic molecular microbiology. *Reviews in Medical Microbiology* 2005;16:59–67.
4. Wadowsky R, Laus S, Libert T, States SJ, Ehrlich GD. Inhibition of PCR-based assay for *Bordetella pertussis* by using calcium alginate fiber and aluminum shaft components of a nasopharyngeal swab. *Journal of Clinical Microbiology* 1994;32:1054–1057.
5. Arbefeville S, Ferrieri P. Comparison of rates of positivity for *Bordetella pertussis* by real-time PCR between specimens collected with rayon swabs on aluminum wire shaft in Amies gel with charcoal and specimens collected with flocked swabs in universal viral transport medium during an epidemic. *Journal of Clinical Microbiology* 2014;52:2656–2658.
6. Dauphin L, Marston CK, Bhullar V, Baker D, Rahman M, Hossain MJ, Chakraborty A, Khan SU, Hoffmaster AR. Swab protocol for rapid laboratory diagnosis of cutaneous anthrax. *Journal of Clinical Microbiology* 2012;50:3960–3967.
7. Rafi W, Chandramuki A, Mani R, Satishchandra P, Shankar SK. Rapid diagnosis of acute bacterial meningitis: role of a broad range 16S rRNA polymerase chain reaction. *Emergency Medicine* 2010;38:225–230.
8. Nolte F, Caliendo AM. *Molecular detection and identification of microorganisms*. 8th ed. Washington, DC: ASM Press; 2003.
9. Nolte F. *Molecular diagnostic methods for infectious diseases*. 3rd ed. Wayne, PA: National Committee for Clinical Laboratory Standards; 2015.
10. Palarasah Y, Skjoedt MO, Vitved L, Andersen TE, Skjoedt K, Koch C. Sodium polyanethole sulfonate as an inhibitor of activation of complement function in blood culture systems. *Journal of Clinical Microbiology* 2010;48:908–914.
11. Buckwalter S, Sloan LM, Cunningham SA, Espy MJ, Uhl JR, Jones MF, Vetter EA, Mandrekar J, Cockerill FR 3rd, Pritt BS, Patel R, Wengenack NL. Inhibition controls for qualitative real-time PCR assays: are they necessary for all specimen matrices? *Journal of Clinical Microbiology* 2014;52:2139–2143.
12. Madej R, Davis J, Holden MJ, Kwang S, Labourier E, Schneider GJ. International standards and reference materials for quantitative molecular infectious disease testing. *Journal of Molecular Diagnostics* 2010;12:133–143.
13. Espy M, Uhl JR, Sloan LM, Buckwalter SP, Jones MF, Vetter EA, Yao JD, Wengenack NL, Rosenblatt JE, Cockerill FR, Smith TF. Real-time PCR in clinical microbiology: applications for routine laboratory testing. *Clinical Microbiology Reviews* 2006;19:165–256.
14. Sabat A, van Zanten E, Akkerboom V, Wisselink G, van Slochteren K, de Boer RF, Hendrix R, Friedrich AW, Rossen JWA, Kooistra-Smid AMDM. Targeted next-generation sequencing of the 16S-23S rRNA region for culture-independent bacterial identification—increased discrimination of closely related species. *Scientific Reports* 2017;7:3434.
15. Doern C, Butler-Wu SM. Emerging and future applications of matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry in the clinical microbiology laboratory: a report of the Association for Molecular Pathology. *Journal of Molecular Diagnostics* 2016;18:789–802.
16. Schubert S, Kostrzewa M. MALDI-TOF MS in the microbiology laboratory: current trends. *Current Issues in Microbiology* 2017;23:17–20.
17. Singhal N, Kumar M, Kanaujia PK, Virdi JS. MALDI-TOF mass spectrometry: an emerging technology for microbial identification and diagnosis. *Frontiers in Microbiology* 2015;6:791–800.
18. Marchello C DA, Thai TN, Han DS, Ebell MH. Prevalence of atypical pathogens in patients with cough and community-acquired pneumonia: a meta-analysis. *Family Medicine* 2016;14:552–566.
19. Cole S, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R,

Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature* 1998;393:537–544.

20. Diaz MH, Winchell J. The evolution of advanced molecular diagnostics for the detection and characterization of *Mycoplasma pneumoniae*. *Frontiers in Microbiology* 2016;7:232.

21. Khanna M, Fan J, Pehler-Harrington K, Waters C, Douglass P, Stallock J, Kehl S, Henrickson KJ. The pneumoplex assays, a multiplex PCR-enzyme hybridization assay that allows simultaneous detection of five organisms, *Mycoplasma pneumoniae, Chlamydia (Chlamydophila) pneumoniae, Legionella pneumophila, Legionella micdadei,* and *Bordetella pertussis,* and its real-time counterpart. *Journal of Clinical Microbiology* 2005;43:565–571.

22. Kumar S, Wang L, Fan J, Kraft A, Bose ME, Tiwari S, Van Dyke M, Haigis R, Luo T, Ghosh M, Tang H, Haghnia M, Mather EL, Weisburg WG, Henrickson KJ. Detection of 11 common viral and bacterial pathogens causing community-acquired pneumonia or sepsis in asymptomatic patients by using a multiplex reverse transcription-PCR assay with manual (enzyme hybridization) or automated (electronic microarray) detection. *Journal of Clinical Microbiology* 2008;46:3063–3072.

23. Murdoch DR, Anderson T, Beynon KA, Chua A, Fleming AM, Laing RT, Town GI, Mills GD, Chambers ST, Jennings LC. Evaluation of a PCR assay for detection of *Streptococcus pneumoniae* in respiratory and nonrespiratory samples from adults with community-acquired pneumonia. *Journal of Clinical Microbiology* 2003;41:63–66.

24. Varghese R, Jayaraman R, Veeraraghavan B. Current challenges in the accurate identification of *Streptococcus pneumoniae* and its serogroups/serotypes in the vaccine era. *Journal of Microbiological Methods* 2017;141:48–54.

25. Marín M, Cercenado E, Sánchez-Carrillo C, Ruiz A, Gómez González Á, Rodríguez-Sánchez B, Bouza E. Accurate differentiation of *Streptococcus pneumoniae* from other species within the *Streptococcus mitis* group by peak analysis using MALDI-TOF MS. *Frontiers in Microbiology* 2017;8:698.

26. Harju I, Lange C, Kostrzewa M, Maier T, Rantakokko-Jalava K, Haanperä M. Improved differentiation of *Streptococcus pneumoniae* and other *S. mitis* group Streptococci by MALDI Biotyper using an improved MALDI Biotyper database content and a novel result interpretation algorithm. *Journal of Clinical Microbiology* 2017;55:914–922.

27. Malm K, Andersson S, Fredlund H, Norrgren H, Biague A, Månsson F, Ballard R, Unemo M. Analytical evaluation of nine serological assays for diagnosis of syphilis. *Journal of the European Academy of Dermatology and Venereology* 2015;29:2369–2376.

28. Li L, Cai B, Tao C, Wang L. Performance evaluation of CLIA for *Treponema pallidum* specific antibodies detection in comparison with ELISA. *Journal of Clinical Laboratory Analysis* 2016;30:216–222.

29. Fraser C, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, Bult CJ, Kerlavage AR, Sutton G, Kelley JM, Fritchman RD, Weidman JF, Small KV, Sandusky M, Fuhrmann J, Nguyen D, Utterback TR, Saudek DM, Phillips CA, Merrick JM, Tomb JF, Dougherty BA, Bott KF, Hu PC, Lucier TS, Peterson SN, Smith HO, Hutchison CA, Venter JC. The minimal gene complement of *Mycoplasma genitalium*. *Science* 1995;270: 397–403.

30. McIver C, Rismanto N, Smith C, Naing ZW, Rayner B, Lusk MJ, Konecny P, White PA, Rawlinson WD. Multiplex PCR testing detection of higher-than-expected rates of cervical *Mycoplasma, Ureaplasma,* and *Trichomonas* and viral agent infections in sexually active Australian women. *Journal of Clinical Microbiology* 2009;47:1358–1363.

31. Leli C, Mencacci A, Latino MA, Clerici P, Rassu M, Perito S, Castronari R, Pistoni E, Luciano E, De Maria D, Morazzoni C, Pascarella M, Bozza S, Sensini A. Prevalence of cervical colonization by *Ureaplasma parvum, Ureaplasma urealyticum, Mycoplasma hominis* and *Mycoplasma genitalium* in childbearing age women by a commercially available multiplex real-time PCR: an Italian observational multicentre study. *Journal of Microbiology, Immunology and Infection* 2017;51(2):220–225.

32. LaSala P, Bufton KK, Ismail N, Smith MB. Prospective comparison of R-mix(tm) shell vial system with direct antigen tests and conventional cell culture for respiratory virus detection. *Journal of Clinical Virology* 2007;38:210–216.

33. Leligdowicz A, Feldmann J, Jaye A, Cotten M, Dong T, McMichael A, Whittle H, Rowland-Jones S. Direct relationship between virus load and systemic immune activation in HIV-2 infection. *Journal of Infectious Diseases* 2010;20:114–122.

34. Escudero D, Lurie MN, Mayer KH, King M, Galea S, Friedman SR, Marshall BDL. The risk of HIV transmission at each step of the HIV care continuum among people who inject drugs: a modeling study. *BMC Public Health* 2017;17:614–623.

35. Josko D. Molecular virology in the clinical laboratory. *Clinical Laboratory Science* 2010;23:231–236.

36. Zeh C, Ndiege K, Inzaule S, Achieng R, Williamson J, Chih-Wei Chang J, Ellenberger D, Nkengasong J. Evaluation of the performance of Abbott m2000 and Roche COBAS Ampliprep/COBAS TaqMan assays for HIV-1 viral load determination using dried blood spots and dried plasma spots in Kenya. *PLoS One.* 2017;12(6):e0179316.

37. Parikh U, McCormick K, van Zyl G, Mellors JW. Future technologies for monitoring HIV drug resistance and cure. *Current Opinions in HIV AIDS* 2017;12:182–189.

38. Shen C, Yu X, Harrison RW, Weber IT. Automated prediction of HIV drug resistance from genotype data. *BMC Bioinformatics* 2016;17:278–285.

39. Lapointe H, Dong W, Lee GQ, Bangsberg DR, Martin JN, Mocello AR, Boum Y, Karakas A, Kirkby D, Poon AF, Harrigan PR, Brumme CJ. HIV drug resistance testing by high-multiplex "wide" sequencing on the MiSeq instrument. *Antimicrobial Agents and Chemotherapy* 2015;59:6824–6833.

40. Raab-Traub N, Flynn K. The structure of the termini of the Epstein-Barr virus as a marker of clonal cellular proliferation. *Cell* 1986;47:883–889.

41. Huh H, Park JE, Kim JY, Yun SA, Lee MK, Lee NY, Kim JW, Ki CS. Performance of the Real-Q EBV Quantification Kit for

Epstein-Barr virus DNA quantification in whole blood. *Annals of Laboratory Medicine* 2017;37:147–150.

42. Lyons M, Kunnathur VA, Rouster SD, Hart KW, Sperling MI, Fichtenbaum CJ, Sherman KE. Prevalence of diagnosed and undiagnosed hepatitis C in a Midwestern urban emergency department. *Clinical Infectious Diseases* 2016;62:1066–1071.

43. Graham S. The human papillomavirus replication cycle, and its links to cancer progression: a comprehensive review. *Clinical Science* 2017;131:2201–2221.

44. Ginocchio C, St. George K. Likelihood that an unsubtypeable influenza A result in the Luminex xTAG Respiratory Virus Panel is indicative of novel A/H1N1 (swine-like) influenza. *Journal of Clinical Microbiology* 2009;47:1027–1029.

45. Gao S, Anderson TK, Walia RR, Dorman KS, Janas-Martindale A, Vincent AL. The genomic evolution of H1 influenza A viruses from swine detected in the United States between 2009 and 2016. *Journal of General Virology* 2017;98(8):2001–2010.

46. Abed Y, Carbonneau J, L'Huillier AG, Kaiser L, Boivin G. Droplet digital PCR to investigate quasi-species at codons 119 and 275 of the A(H1N1)pdm09 neuraminidase during zanamivir and oseltamivir therapies. *Journal of Medical Virology* 2017;89:737–741.

47. Gullett J, Nolte FS. Quantitative nucleic acid amplification methods for viral infections. *Clinical Chemistry* 2015;61:72–78.

48. Piao J, Jiang J, Xu B, Wang X, Guan Y, Wu W, Liu L, Zhang Y, Huang X, Wang P, Zhao J, Kang X, Jiang H, Cao Y, Zheng Y, Jiang Y, Li Y, Yang Y, Chen W. Simultaneous detection and identification of enteric viruses by PCR-mass assay. *PLoS One* 2012;7:e2251.

49. Kriegsmann M, Wandernoth P, Lisenko K, Casadonte R, Longuespée R, Arens N, Kriegsmann J. Detection of HPV subtypes by mass spectrometry in FFPE tissue specimens: a reliable tool for routine diagnostics. *Journal of Clinical Pathology* 2017;70:417–423.

50. Downard K. Proteotyping for the rapid identification of influenza virus and other biopathogens. *Chemical Society Reviews* 2013;42:8584–8598.

51. Frickmann H, Zautner AE, Moter A, Kikhney J, Hagen RM, Stender H, Poppert S. Fluorescence in situ hybridization (FISH) in the microbiological diagnostic routine laboratory: a review. *Critical Reviews in Microbiology* 2017;43:263–293.

52. Fenselau C, Demirev PA. Characterization of intact microorganisms by MALDI mass spectrometry. *Mass Spectrometry Reviews* 2001;20:157–171.

53. Idelevich E, Grünastel B, Becker K. Rapid detection and identification of Candidemia by direct blood culturing on solid medium by use of lysis-centrifugation method combined with matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS). *Journal of Clinical Microbiology* 2016;55:97–100.

54. Moreira OC, Verly T, Finamore-Araujo P, Gomes SAO, Lopes CM, de Sousa DM, Azevedo LR, da Mota FF, d'Avila-Levy CM, Santos-Mallet JR, Britto C. Development of conventional and real-time multiplex PCR-based assays for estimation of natural infection rates and *Trypanosoma cruzi* load in triatomine vectors. *Parasites & Vectors* 2017;10:404.

55. Weigel L, Clewell DB, Gill SR, Clark NC, McDougal LK, Flannagan SE, Kolonay JF, Shetty J, Killgore GE, Tenover FC. Genetic analysis of a high-level vancomycin-resistant isolate of *Staphylococcus aureus*. *Science* 2003;302:1569–1571.

56. Unissa A, Subbian S, Hanna LE, Selvakumar N. Overview on mechanisms of isoniazid action and resistance in *Mycobacterium tuberculosis*. *Infection, Genetics and Evolution* 2016;45:474–492.

57. Metan G, Zarakolu P, Unal S. Rapid detection of antibacterial resistance in emerging gram-positive cocci. *Journal of Hospital Infection* 2005;61:93–99.

58. Kelly A, Mathema B, Larson EL. Carbapenem-resistant Enterobacteriaceae in the community: a scoping review. *International Journal of Antimicrobial Agents* 2017;50:127–134.

59. Ceyssens P, Soetaert K, Timke M, Van den Bossche A, Sparbier K, De Cremer K, Kostrzewa M, Hendrickx M, Mathys V. Matrix-assisted laser desorption ionization-time of flight mass spectrometry for combined species identification and drug sensitivity testing in mycobacteria. *Journal of Clinical Microbiology* 2017;55:624–634.

60. Oviaño M, Sparbier K, Barba MJ, Kostrzewa M, Bou G. Universal protocol for the rapid automated detection of carbapenem-resistant Gram-negative bacilli directly from blood cultures by matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF/MS). *International Journal of Antimicrobial Agents* 2016; 48:655–660.

61. Hoyos-Mallecot Y, Riazzo C, Miranda-Casas C, Rojo-Martín MD, Gutiérrez-Fernández J, Navarro-Marí JM. Rapid detection and identification of strains carrying carbapenemases directly from positive blood cultures using MALDI-TOF MS. *Journal of Microbiological Methods* 2014;105:98–101.

62. Leclercq R, Derlot E, Duval J, Courvalin P. Plasmid-mediated resistance to vancomycin and teicoplanin in *Enterococcus faecium*. *New England Journal of Medicine* 1988;319:157–161.

63. van Belkum A, Welker M, Pincus D, Charrier JP, Girard V. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry in clinical microbiology: what are the current issues? *Annals of Laboratory Medicine* 2017;37:475–483.

64. Meacci F, Orru G, Iona E, Giannoni F, Piersimoni C, Pozzi G, Fattorini L, Oggioni MR. Drug resistance evolution of a *Mycobacterium tuberculosis* strain from a noncompliant patient. *Journal of Clinical Microbiology* 2005;43:3114–3120.

65. Fournier P, Drancourt M, Raoult D. New laboratory tools for emerging bacterial challenges. *Clinical Infectious Diseases* 2017;65:S39–S49.

66. Lee F. Diagnostics and laboratory role in outbreaks. *Current Opinion in Infectious Diseases* 2017;30:419–442.

67. Tenover F, Arbeit RD, Goering RV, Mickelsen PA, Murray BE, Persing DH, Swaminathan B. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. *Journal of Clinical Microbiology* 1995;33:2233–2239.

68. Stingu C, Eschrich K, Thiel J, Borgmann T, Schaumann R, Rodloff AC. Identification of viridans streptococci with matrix-assisted laser desorption; ionization time-of-flight mass spectrometry by an in-house method and a commercially available system. *Annals of Laboratory Medicine* 2017;37:434–437.

69. Valentine N, Wunschel S, Wunschel D, Petersen C, Wahl K. Effect of culture conditions on microorganism identification by matrix-assisted laser desorption ionization mass spectrometry. *Applied Environmental Microbiology* 2005;71:58–64.

70. Carbonnelle E, Beretti JL, Cottyn S, Quesne G, Berche P, Nassif X, Ferroni A. Rapid identification of Staphylococci isolated in clinical microbiology laboratories by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Journal of Clinical Microbiology* 2007;45:2156–2161.

71. Donohue M, Best JM, Smallwood AW, Kostich M, Rodgers M, Shoemaker JA. Differentiation of Aeromonas isolated from drinking water distribution systems using matrix-assisted laser desorption/ionization-mass spectrometry. *Analytical Chemistry* 2007;79:1939–1946.

72. Huang X, Liu Q, Huang X, Nie Z, Ruan T, Du Y, Jiang G. Fluorographene as a mass spectrometry probe for high-throughput identification and screening of emerging chemical contaminants in complex samples. *Analytical Chemistry* 2016;89:1307–1314.

73. Pholwat S, Sakai F, Turner P, Vidal JE, Houpt ER. Development of a TaqMan array card FOR Pneumococcal serotyping on isolates and nasopharyngeal samples. *Journal of Clinical Microbiology* 2016;54:1842–1850.

74. Driscoll A, Karron RA, Bhat N, Thumar B, Kodani M, Fields BS, Whitney CG, Levine OS, O'Brien KL, Murdoch DR. Evaluation of fast-track diagnostics and TaqMan array card real-time PCR assays for the detection of respiratory pathogens. *Journal of Microbiological Methods* 2014;107:222–226.

75. Opota O, Brouillet R, Greub G, Jaton K. Methods for real-time PCR-based diagnosis of *Chlamydia pneumoniae, Chlamydia psittaci,* and *Chlamydia abortus* infections in an opened molecular diagnostic platform. *Methods in Molecular Biology* 2017;1616:171–181.

76. Olive D, Bean P. Principles and applications of methods for DNA-based typing of microbial organisms. *Journal of Clinical Microbiology* 1999;37:1661–1669.

Chapter 12

Molecular Detection of Inherited Diseases

Outline

- THE MOLECULAR BASIS OF INHERITED DISEASES
- CHROMOSOMAL ABNORMALITIES
- PATTERNS OF INHERITANCE IN SINGLE-GENE DISORDERS
- MOLECULAR BASIS OF SINGLE-GENE DISORDERS
    - *Lysosomal Storage Diseases*
    - *Factor V Leiden*
    - *Prothrombin*
    - *Methylenetetrahydrofolate Reductase*
    - *Hemochromatosis*
    - *Cystic Fibrosis*
    - *Cytochrome P-450*
- SINGLE-GENE DISORDERS WITH NONCLASSICAL PATTERNS OF INHERITANCE
    - *Mutations in Mitochondrial Genes*
    - *Nucleotide-Repeat Expansion Disorders*
        - Fragile X Syndrome
        - Huntington Disease
        - Idiopathic Congenital Central Hypoventilation Syndrome
        - Other Nucleotide Expansion Disorders
    - *Genomic Imprinting*
    - *Multifactorial Inheritance*
- LIMITATIONS OF MOLECULAR TESTING

Objectives

12.1 Describe Mendelian patterns of inheritance as exhibited by family pedigrees.

12.2 Illustrate abnormalities in chromosome number and structure.

12.3 Define penetrance and variable expressivity.

12.4 Relate disease syndromes with affected genes.

12.5 Give examples of laboratory methods designed to detect single-gene disorders.

12.6 Discuss non-Mendelian inheritance, and give examples of these types of inheritance, such as mitochondrial disorders and nucleotide-repeat expansion diseases.

12.7 Show how genomic imprinting can affect disease phenotype.

Genetic and cytogenetic analyses are a critical component of diagnostic testing, especially for diseases that arise from known genetic events. The identification of a molecular or chromosomal abnormality is a direct observation of the source of some diseases. This chapter presents examples of clinical laboratory tests commonly performed in molecular genetics using these techniques.

THE MOLECULAR BASIS OF INHERITED DISEASES

Mutations are changes in DNA nucleotide sequences. These changes range from single base pair or point mutations of various types to chromosomal. Not all mutations lead to disease.

Polymorphisms are proportionately represented genotypes in a given population. Sequence polymorphisms can be located within genes or outside of genes. Benign polymorphisms are useful for mapping disease genes, determining parentage, and identity testing. Balanced polymorphisms can have offsetting phenotypes.

Epigenetic alterations do not change the primary DNA sequence. Epigenetic changes consist of three different forms: DNA methylation, usually alterations of cytosine in CpG islands; **genomic imprinting;** and chromatin remodeling. DNA methylation mostly down-regulates RNA transcription. Genomic imprinting selectively inactivates chromosomal regions (e.g., X chromosome inactivation). Chromatin remodeling sequesters large regions of chromosomal DNA through protein binding and histone modification. Histone modification controls the availability of DNA for RNA transcription.

Mutations in germ cells result in inherited disease. Mutations in somatic cells result in cancer and some congenital malformations. Diseases with genetic components are often referred to as **congenital** ("born with") diseases. Congenital disorders are not necessarily heritable, however. Congenital disorders are those present in individuals at birth. Specifically, congenital disorders result when some factor, such as a drug, a chemical, an infection, or an injury, upsets the developmental process. Thus, a baby can have a heritable disease, such as hemophilia, that can be passed on to future generations or a congenital condition, such as spina bifida, that cannot be passed to offspring.

Advanced Concepts

According to the **Lyon hypothesis** (or Lyon's hypothesis) first stated by Mary Lyon in 1961, only one X chromosome remains genetically active in females.[1,2] In humans, one X chromosome is inactivated at random about the 16th day of embryonic development. The inactive X can be seen as a **Barr body** (X chromatin) in the interphase nucleus. Not all X genes are shut off in the inactivated X chromosome. Furthermore, reactivation of genes on the inactivated X occurs in germ cells before the first meiotic division for production of eggs.

CHROMOSOMAL ABNORMALITIES

Genome mutations (abnormalities in chromosome number) can be detected by karyotyping, ploidy analysis by flow cytometry, and fluorescent in situ hybridization (FISH). **Polyploidy** (more than two of any autosome) in animals usually results in infertility and abnormal appearance. Aneuploidy (gain or loss of any autosome) occurs with 0.5% frequency in term pregnancies and 50% in spontaneous abortions. Aneuploidy is caused by erroneous separation of chromosomes during egg or sperm production (chromosomal non-disjunction). Autosomal trisomy/monosomy (three copies/one copy of a chromosome instead of two) results from fertilization of gametes containing an extra chromosome or missing a chromosome ($n + 1$ or $n - 1$ gametes, respectively). Autosomal monosomy is generally, but not always, incompatible with life. Sex chromosome aneuploidy is more frequently tolerated, although it is associated with phenotypic abnormalities.

Mosaicism, two or more genetically distinct populations of cells from one zygote in an individual (in contrast to chimerism: two or more genetically distinct cell populations from different zygotes in an individual), results from mutation events affecting somatic or germ cells. Early segregation errors during fertilized egg division occasionally give rise to mosaicism. Mosaicism is relatively common with sex chromosomes; for example,

TABLE 12.1 Examples of Genome Mutations

Disorder	Genetic Abnormality	Incidence	Clinical Features
Down syndrome	Trisomy 21, 47,XY,+21	1/700 live births	Flat facial profile, mental retardation, cardiac problems, risk of acute leukemia, eventual neuropathological disorders, abnormal immune system
Edward syndrome	Trisomy 18, 47,XY,+18	1/3,000 live births	Severe, clenched fist; survival less than 1 year
Patau syndrome	Trisomy 13, 47,XY,+13	1/5,000 live births	Cleft palate, heart damage, mental retardation, survival usually less than 6 mo
Klinefelter syndrome	47,XXY	1/850 live births	Male hypogonadism, long legs, gynecomastia (male breast enlargement), low testosterone level
XYY syndrome	47,XYY	1/1,000 live births	Excessive height, acne, 1%–2% behavioral disorders
Turner syndrome	45,X and variants	1/2,000 live births	Bilateral neck webbing, heart disease, failure to develop secondary sex characteristics, hypothyroidism
Multi X females	47,XXX; 48,XXXX	1/1,200 newborn females	Mental retardation increases with increasing X

45,X/47,XXX (normal female chromosome complement is 46,XX). In this case, later **nondisjunction** will yield additional populations. Rarely, autosomal haploids will be lost with the retention of the **triploid** lineage (e.g., 45,XY,-21, 46,XY/47,XY,+21 → 46,XY/47,XY,+21). Examples of genome mutations are shown in Table 12.1 Chromosome mutations (abnormalities in chromosome structure) larger than 4 million base pairs (bp) can be seen by karyotyping; smaller irregularities can be seen with the higher resolution of FISH or microarray technology. Structural alterations include translocations (reciprocal, nonreciprocal), inversions (paracentric, pericentric), deletions (terminal, interstitial, ring), duplications (isochromosomes), **marker chromosomes,** and derivative chromosomes.

Structural mutations require breakage and reunion of DNA. Chromosomal breakage is caused by chemicals and radiation. Chromosomal breakage also results from chromosome breakage syndromes (e.g., Fanconi anemia, Bloom syndrome, and ataxia telangiectasia). Some aberrations have no immediate phenotypic effect (reciprocal translocations, inversions, some deletions, some insertions). Others can be deleterious to cells including lethality. Chromosome translocations are usually somatic events (not inherited) and are most commonly seen in cancer. Approximately 7.4% of conceptions have chromosome mutations. Chromosome mutations are observed in 50% of spontaneous abortions and 5% of stillbirths. Examples of diseases arising from inherited chromosome structure abnormalities are shown in Table 12.2.

PATTERNS OF INHERITANCE IN SINGLE-GENE DISORDERS

Most phenotypes result from the interaction of multiple genetic and environmental factors. Some phenotypes, however, are caused by alteration of a single gene. If the phenotype occurs as predicted by Mendelian genetics, patterns of inheritance can be established. Patterns of inheritance **(transmission patterns)** are determined by examination of family histories. A **pedigree** is a diagram of the inheritance pattern of a phenotype of family members (Fig 12.1). There are three main Mendelian transmission patterns: autosomal dominant, autosomal recessive, and **X-linked** or sex-linked recessive. These patterns refer to the disease phenotype.

TABLE 12.2 Examples of Chromosomal Mutations

Disorder	Genetic Abnormality	Incidence	Clinical Features
DiGeorge syndrome and velocardiofacial syndrome	del(22q)	1/4,000 live births	CATCH 22 (cardiac abnormality/abnormal facies, T-cell deficit, cleft palate, hypercalcemia)
Cri du chat syndrome	del(5p)	1/20,000–1/50,000 live births	Growth deficiency, catlike cry in infancy, small head, mental retardation
Contiguous gene syndrome; Wilms' tumor, aniridia, genitourinary anomalies, mental retardation syndrome	del(11p)	1/15,000 live births	Aniridia (absence of iris), hemihypertrophy (one side of the body seems to grow faster than the other), and other congenital anomalies

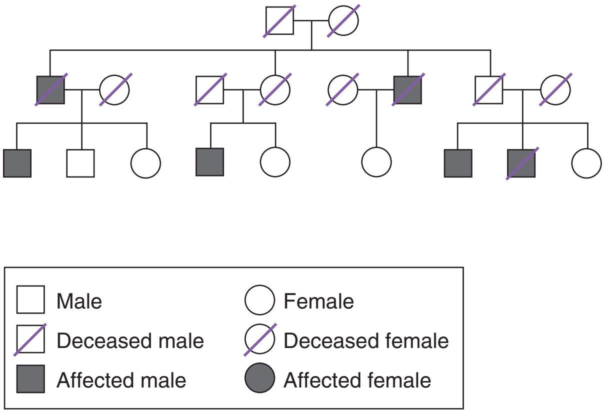


FIGURE 12.1 A pedigree is a diagram of family phenotype or genotype. The pedigree will display the transmission pattern of a disease. Phenotypic traits are followed by coloring or patterns in the symbols. Lack of the trait is indicated by open symbols with no coloring.

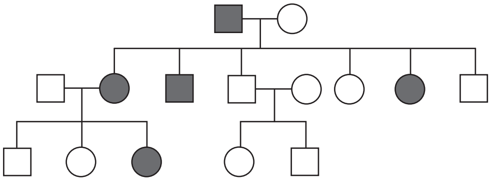


FIGURE 12.2 In an autosomal-dominant transmission pattern, heterozygous individuals express the affected phenotype (filled symbols).

In **autosomal-dominant** transmission, a child of an affected individual and an unaffected mate has a 50% to 100% risk or likelihood of expressing the disease phenotype (Fig. 12.2). **Gain-of-function** mutations are usually dominant because the mutated allele produces sufficient abnormal factor to bring about the affected condition. In complete dominance, the heterozygous phenotype of the offspring is the same as that of the homozygous parent. In partial dominance, the offspring phenotype is variably intermediate between the homozygous and heterozygous parental phenotypes. The parental phenotypes can reappear in the next generation, showing that Mendelian inheritance is still present and that the partial dominance is a manifestation of how the genes are expressed. **Codominant offspring** simultaneously demonstrate the phenotype of both parents. A familiar example of codominance is the ABO blood types. **Dominant-negative** phenotypes are seen in cases of multimeric proteins such as the tumor-suppressor tetramer, p53 (Fig. 12.3). Even though only one allele is mutated, the mutated protein can interfere with the function of the tetramer, producing an abnormal phenotype.

The phenotype of a **loss-of-function** mutation is usually recessive, but it depends on the type of protein affected. Complex metabolic pathways are susceptible to loss-of-function mutations because of extensive interactions between and among proteins. Key structural proteins, especially multimeric complexes, risk dominant negative phenotypes. Gain-of-function mutations are less common than loss-of-function mutations. Gain-of-function mutations include gene-expression/stability defects that generate gene products at inappropriate sites or times.

Autosomal recessive is the largest category of Mendelian disorders. The recurrence risk is 25% if siblings are affected, indicating the presence of the recessive mutation in both of the parents. The "margin of safety," that is, having two copies of every gene, requires the loss of the normal allele through somatic events (loss of heterozygosity) or homozygosity for the manifestation of the recessive disease phenotype. Autosomal-recessive diseases are more often observed as a result of two individuals heterozygous for the same mutation producing offspring (Fig. 12.4). New mutations are rarely detected in autosomal-recessive transmission patterns. Inborn errors of metabolism are usually autosomal recessive. Risk factors for neoplastic diseases also fall in this category. Polygenic disorders (caused by multiple gene mutations) such as type 2 diabetes may also be inborn errors of metabolism.[3]

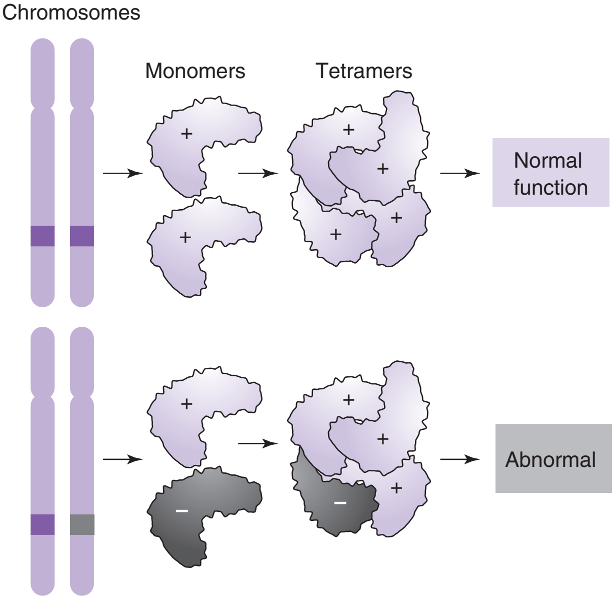


FIGURE 12.3 Dominant negative mutations affect multimeric proteins. In this illustration, a single mutant monomer affects the function of the assembled tetramer.

Advanced Concepts

Autosomal-dominant mutations can originate from **gonadal mosaicism** of **new mutations** in germ cells, that is, DNA changes that arise in cells that produce eggs or sperm. Establishment of a new mutation as a dominant mutation in a family or in a population is influenced by its effect on reproductive fitness.

Almost all **sex-linked** disorders are X-linked because relatively few genes are carried on the Y chromosome. X-linked mutations are almost always recessive, but there are X-linked dominant diseases (e.g., vitamin D–resistant rickets). Even though one X chromosome is inactivated in females, the inactivation is reversible so that a second copy of X-linked genes is available. In contrast, males are **hemizygous** for X-linked genes, having only one copy on the X chromosome. Males, therefore, are more likely to manifest the disease phenotype (Fig. 12.5).

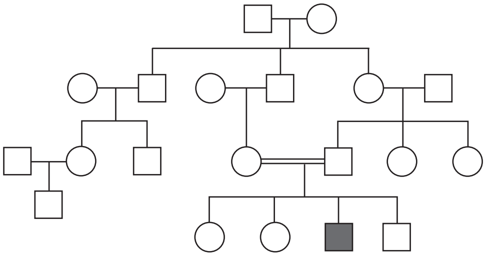


FIGURE 12.4 Autosomal-recessive mutations are not expressed in heterozygotes. The phenotype is displayed only in a homozygous individual; in this illustration, produced by the inbreeding of two cousins (double horizontal line).

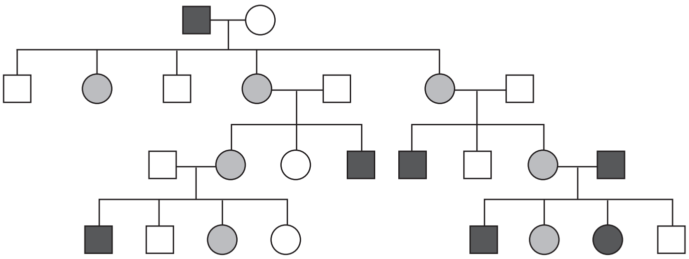


FIGURE 12.5 X-linked recessive diseases are carried by females but manifested most often in males.

Due to the multifactorial nature of most diseases, the same gene mutations are not always manifested in a disease phenotype. **Penetrance** is the frequency of expression of disease phenotype in individuals with a gene lesion. Complete penetrance is the expression of the disease phenotype in every individual with the mutated gene. Complete penetrance is common in homozygous recessive phenotypes. **Variable expressivity** is a range of phenotypes in individuals with the same gene lesion. Variable expressivity also reflects the interaction of other gene products and the environment on the disease phenotype.

MOLECULAR BASIS OF SINGLE-GENE DISORDERS

Single-gene disorders affect structural proteins, cell surface receptor proteins, growth regulators, and enzymes. Examples of diseases resulting from such disorders are shown in Table 12.3. Examples of molecular methods that have been or could be used to detect these gene lesions are also listed. Not all of these methods are in common use in molecular diagnostics. Some diseases are effectively analyzed by morphological studies or clinical chemistry. For instance, hemoglobin S is classically detected by protein electrophoresis. Final diagnosis requires physiological, morphological, and laboratory results.

Lysosomal Storage Diseases

Lysosomes are subcellular organelles in which products of cellular ingestion are degraded by acid hydrolase enzymes (Fig. 12.6). These enzymes work in an acid environment. Substrates come from intracellular turnover **(autophagy)** or outside the cell through phagocytosis or endocytosis **(heterophagy).** Lysosomal storage disorders result from incompletely digested macromolecules due to loss of enzymatic degradation. Storage disorders include defects in proteins required for normal lysosomal function, giving rise to physical abnormalities. The organs affected depend on the location and site of degradation of the substrate material. Examples of storage diseases are shown in Table 12.4. These disorders are screened by gene product testing, that is, measuring the ability of serum enzymes to digest test substrates. With the discovery of genes that code for the enzymes and their subunits, molecular testing has been used to some extent. Mutations can be detected by direct sequencing, usually after an initial biochemical screening test for loss of enzyme activity.

Factor V Leiden

Mutations that lead to abnormal but survival states can be relatively frequently encountered in a population. An example is the hypercoagulation phenotype resulting in mutations in the factor V gene. Discovered in 1994, the Leiden mutation (1691 A→G, R506Q) in the coagulation factor V gene *F5* (1q23) causes a hypercoagulable (thrombophilic) phenotype. This genotype is present in heterozygous form in 4% to 8% of the general population, and 0.06% to 0.25% of the population is homozygous for this mutation. A blood clot or deep venous thrombosis is treated with anticoagulants. The risk of thrombosis increases with contraceptive use in women (Table 12.5).

Several approaches have been taken to test for the Leiden mutation. Polymerase chain reaction (PCR) methods include the use of restriction fragment length polymorphism (PCR-RFLP) or PCR with sequence-specific primers (SSP-PCR; Figs. 12.7 and 12.8). Nonamplification molecular methods, such as Invader technology, have also been developed to test for this gene mutation. Clot-based methods may be used to directly demonstrate thrombophilia before genetic testing. Family history may also be considered.

Prothrombin

Prothrombin is the precursor to thrombin in the coagulation cascade and is required for the conversion of fibrinogen to fibrin. A mutation in the 3′ untranslated region of the gene that codes for prothrombin or coagulation factor II, *F2* (11p11-q12), results in an autosomal-dominant increased risk of thrombosis (see Table 12.5). Laboratories may test for both *F2* and *F5* mutations. Both may be present in the same individual, in which case the risk of thrombosis is greater than with one of the mutations alone. An example of a multiplex PCR-RFLP method to simultaneously test for both mutations was described previously in Chapter 8. In this method, primers that amplify prothrombin and factor V

TABLE 12.3 Single-Gene Disorders and Molecular Methods

Type of Protein	Type of Disease	Example	Gene (Location)	Type of Mutation	Examples of Molecular Methods
Structural	Hemoglobinopathies	Sickle cell anemia	Hemoglobin beta (11p15.5)	Missense	Sequencing, PCR-RFLP
	Connective tissue disorders	Marfan syndrome	Fibrillin (15q21.1)	Missense	Sequencing, linkage analysis
	Cell membrane–associated protein dysfunction	Muscular dystrophy	Dystrophin, DMD (Xp12.2)	Deletion	Southern blot RFLP,[21] multiplex PCR, linkage analysis
Cell surface receptor proteins	Hypercholesteremia	Familial hypercholesteremia	Low-density lipoprotein receptor (19p13.2)	Deletions, point mutations	Probe amplification, sequencing
	Nutritional disorders	Vitamin D–resistant rickets	Vitamin D receptor (12q12-q14)	Point mutations	Southern blot RFLP, sequencing
Cell growth regulators	Fibromas	Neurofibromatosis type 1 (von Recklinghausen disease)	Neurofibromin tumor suppressor (17q11.2)	Missense, frameshift, splice site mutations	Sequencing, linkage analysis
	Fibromas	Neurofibromatosis type 2 (von Recklinghausen disease)	Merlin tumor suppressor, NF-2 (22q12)	Nonsense, frameshift, splice site mutations	Linkage analysis
	Cancer predisposition	Li-Fraumeni syndrome (LFS)*	p53 tumor-suppressor gene, *TP53* (17p13)	Missense mutations	Sequencing, SSCP, DGGE
Enzymes	Metabolic diseases	Alkaptonuria (ochronosis)	Homogentisic acid oxidase (3q21–q23)	Missense, frameshift, splice site mutations	cDNA sequencing, SSCP
	Phenylketonuria	Phenylalanine hydroxylase, PAH or PKU1 (12q24.1)	Splice site, missense mutations, deletions	Ligase chain reaction,[25,26] direct sequencing	
	Immunodeficiencies	Severe combined immunodeficiency	Adenosine deaminase (20q13.11)	Point mutations	Direct sequencing, capillary electrophoresis

*A significant proportion of LFS and LFL (Li-Fraumeni–like) kindred do not have demonstrable *TP53* mutations.

are designed to destroy or produce *Hind*III restriction sites in the presence of the *F5* or *F2* mutation, respectively. The sizes of the amplicons and their restriction fragments allow resolution of both simultaneously by agarose gel electrophoresis. The fragment patterns in each lane reveal the *F2* and *F5* normal or mutant genotypes for each specimen.[4]

Thrombin time, prothrombin time, platelet count, and complete blood count are phenotypic tests that may be performed in addition or prior to molecular analysis.

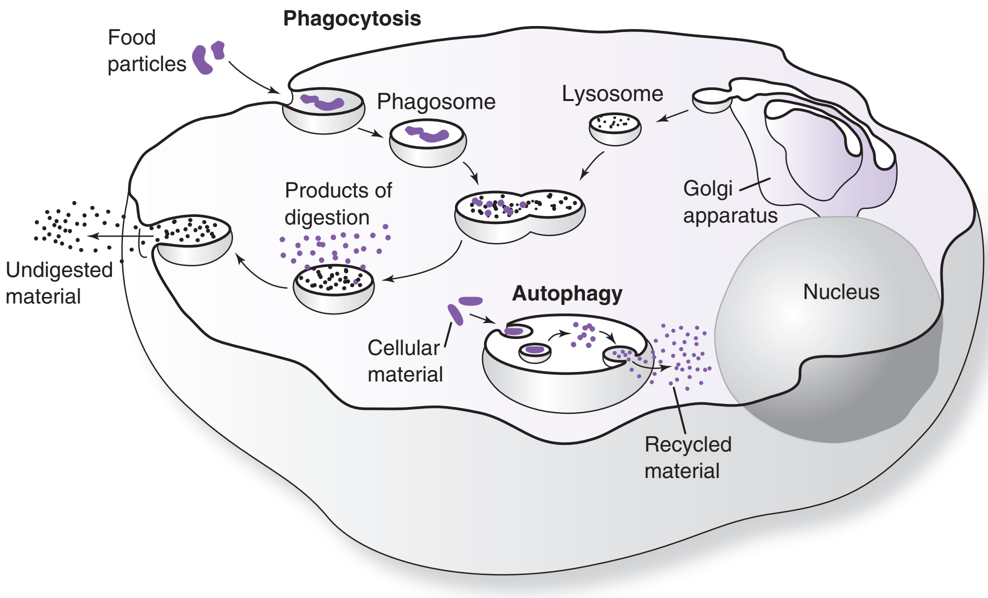

Phagocytosis
Food particles
Phagosome
Lysosome
Golgi apparatus
Products of digestion
Undigested material
Autophagy
Nucleus
Cellular material
Recycled material

FIGURE 12.6 The lysosome is a depository for cell debris. The lysosome contains enzymes that are active in its acid environment to digest proteins delivered from phagocytosis of foreign bodies, endocytosis, and autophagy of internal cellular components such as mitochondria.

TABLE 12.4 Storage Diseases

Substrate Accumulated	Disease
Sphingolipids	Tay–Sachs disease
Glycogen	Von Gierke, McArdle, and Pompe disease
Mucopolysaccharides	Hurler, Sheie (MPS I), Hunter (MPS II), Sanfilippo (MPS III), Morquio (MPS IV), Maroteaux–Lamy (MPS VI), Sly (MPS VII)
Mucolipids	Pseudo-Hurler polydystrophy
Sulfatides	Niemann–Pick disease
Glucocerebrosides	Gaucher disease

TABLE 12.5 Risk of Thrombosis Relative to Normal (1) Under the Indicated Genetic (F5, Prothrombin) and Environmental (OCP) Influences

Status	Risk of Thrombosis
Normal	1
Oral contraceptive (OCP) use	4
Prothrombin mutation, heterozygous	3
Prothrombin mutation + OCP	16
R506Q heterozygous	5–7
R506Q heterozygous + OCP	30–35
R506Q homozygous	80
R506Q homozygous + OCP	100+

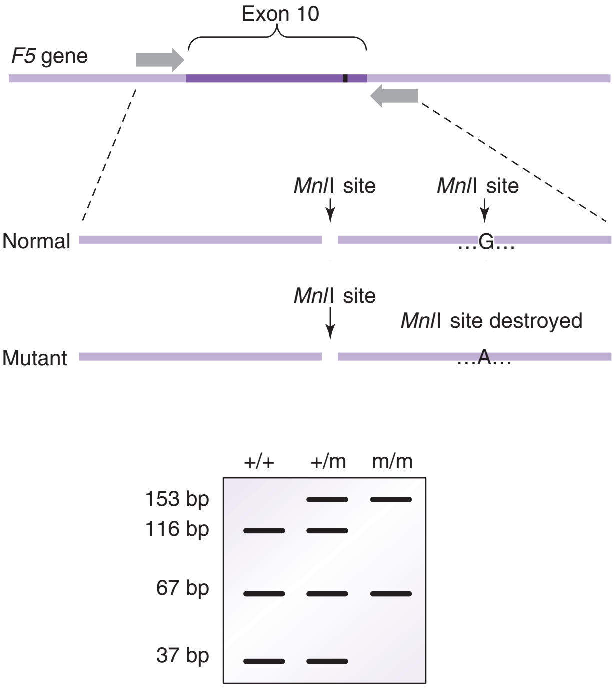


FIGURE 12.7 PCR-RFLP for the factor V Leiden mutation. The R506Q amino acid substitution is caused by a G to A change in exon 10 of the *F5* gene. This DNA mutation destroys an *Mnl*I restriction enzyme site. An amplicon including the site of the mutation, when cut with *Mnl*I, will yield three fragments in normal DNA (+/+) and two products in homozygous mutant DNA (m/m). A heterozygous specimen (+/m) will yield a combination of the normal and mutant pattern.

Automated systems that measure changes in light transmittance during clot formation generating a curve for mathematical analysis replace some of the manual methods. To further identify the genetic cause of abnormal coagulation, sequencing of factors IX and XIII is performed in addition to factor V and II analysis.

Methylenetetrahydrofolate Reductase

Deficiency of the 5,10-methylenetetrahydrofolate reductase *(MTHFR)* gene product is an autosomal-recessive disorder that results in increased homocysteine levels (hyperhomocysteinemia), causing a predisposition to venous and arterial thrombosis.

MTHFR catalyzes the conversion of 5,10-methylenetetrahydrofolate to 5-methyltetrahydrofolate, a co-substrate for conversion of homocysteine to methionine (Fig. 12.9). Two (of more than a dozen) genetic polymorphisms, 677C>T (p.A222V) and 1298 A>C (p.E429A), are associated with deficiencies in folate metabolism. These variants are detectable by standard or multiplex PCR with RFLP using restriction enzymes *Hinf*I and *Mbo*II or sequencing. Multiplex qPCR and high-resolution melt-curve methods have also been developed.[5,6]

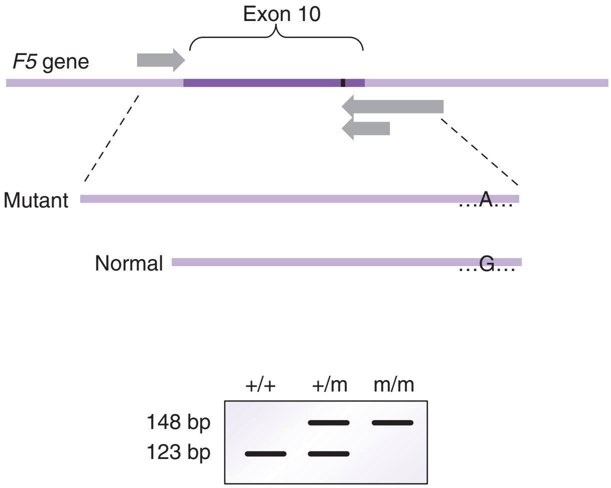


FIGURE 12.8 In sequence-specific PCR, a primer with thymidine as its final 3′ base will yield a product only if the adenine nucleotide is present. The resulting 148 bp PCR product reflects the presence of the mutation. By designing a primer slightly shorter than but complementary to the normal (G) in the template, a distinct, shorter 123-bp normal product is amplified. In a heterozygous individual, both products will appear.

Hemochromatosis

Hemochromatosis is an autosomal-recessive condition that causes overabsorption of iron from food. Iron accumulation subsequently causes pancreas, liver, and skin damage; heart disease; and diabetes. Classically, diagnosis is made through measurement of blood iron levels, transferrin saturation, or liver biopsy.

At the molecular level, hemochromatosis is caused by dysfunction of the hemochromatosis type I *HFE* or *HLA-H* gene product. *HFE* (6p21.3) codes for a membrane-bound protein that binds with β_2 microglobulin

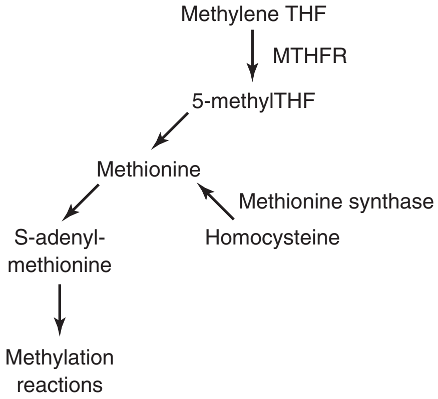


FIGURE 12.9 MTHFR catalyzes the conversion of 5,10-methylenetetrahydrofolate to 5-methyltetrahydrofolate, a co-substrate for homocysteine-mediated remethylation to methionine. Mutations in this gene are associated with thrombophilia and methylenetetrahydrofolate reductase deficiency.

and transferrin on the membrane of cells in the small intestine and also on the placenta. The protein directs iron absorption based on cellular iron loads. In the absence of HFE function, intestinal cells do not sense iron stores, and iron absorption continues into overload.

The most frequently observed mutation in hemochromatosis is C282Y, found in approximately 10% of the Caucasian population, with a disease frequency of 2 to 3 per 1,000 people. Other mutations most frequently detected in the HFE protein are H63D and S65C (Fig. 12.10). Clinical symptoms and increased serum ferritin and transferrin-iron saturation are indications for mutation testing. The C282Y mutation is detectable using PCR-RFLP (Fig. 12.11).

Cystic Fibrosis

Cystic fibrosis (CF) is a life-threatening autosomal-recessive disorder that causes severe lung damage and nutritional deficiencies. With earlier detection by genetic analysis and improved treatment strategies, people with CF can live more comfortably surviving beyond the fourth decade of life. CF affects the cells that produce mucus, sweat, saliva, and digestive juices. Normally, these secretions are thin, but in CF, a defective gene causes the secretions to become thick and sticky. Respiratory failure is the most dangerous consequence of CF.

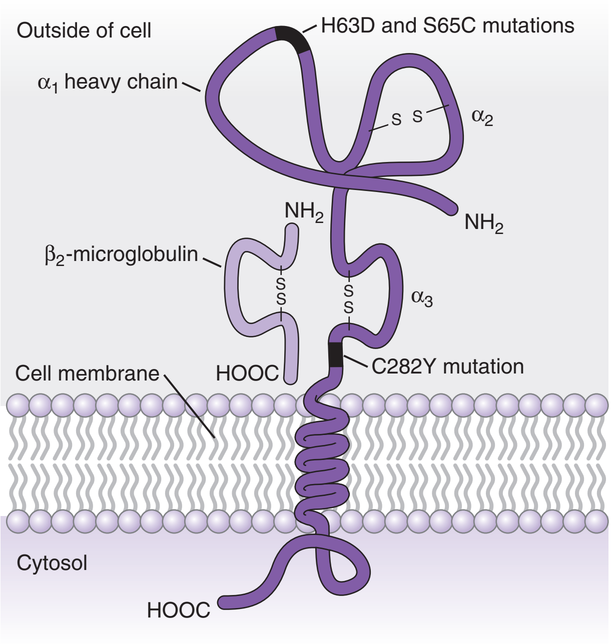


FIGURE 12.10 The HFE protein is associated with β_2-microglobulin in the cell membrane. The location of the frequently occurring mutations is shown.

CF is caused by loss of function of the CF transmembrane conductance regulator, the *CFTR* gene (7q31.2). The gene codes for a chloride channel membrane protein (Fig. 12.12). The first and most frequently observed mutation in *CFTR* is a 3-bp deletion that removes a phenylalanine residue from position 508 of the protein (F508del).[7] More than 1,900 other mutations and variations have been reported in and around the *CFTR* gene in diverse populations. In a European survey of over 25,000 patients, the seven most frequent mutations were F508del, G542X, G551D, N1303K, R117H, W1282X, and 1717-1G->A. These variants accounted for 75% of the alleles, with F508del present in 66.7% of the samples.[8] A list of mutations, their locations, and references is available through the Human Genome Variation Society at http://www.genet.sickkids.on.ca.

Genetic testing for CF is important for diagnosis and genetic counseling because early intervention is most effective in relieving symptoms of the disease. Molecular tests have been designed to detect a variety of mutations that have been described in CF[9] and include

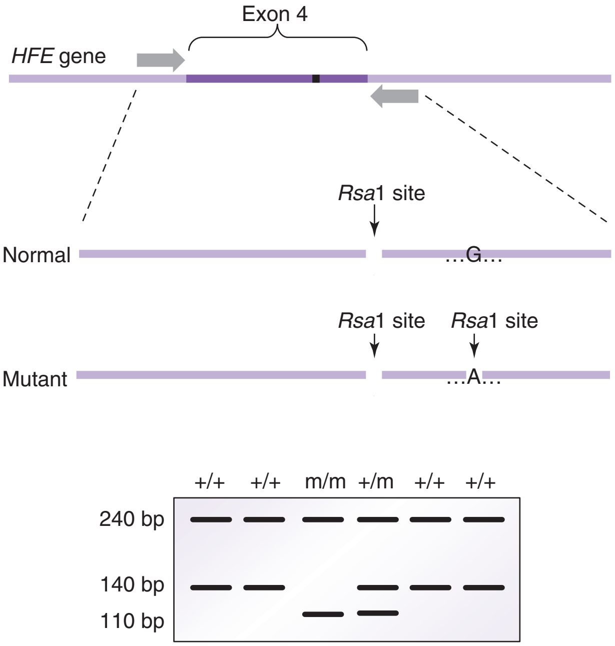

Exon 4
HFE gene
*Rsa*1 site
Normal
...G...
*Rsa*1 site
*Rsa*1 site
Mutant
...A...
+/+
+/+
m/m
+/m
+/+
+/+
240 bp
140 bp
110 bp

FIGURE 12.11 Detection of the C282Y mutation by PCR-RFLP. The G→A mutation in exon 4 of the *HFE* gene produces a site for the restriction enzyme, *Rsa*1. This region is first amplified using primers flanking exon 4 of the gene (arrows). In a normal specimen, the enzyme will produce two fragments, 240 bp and 140 bp. If the mutation is present, the 140-bp normal fragment is cut to a 110-bp and a 30-bp fragment (the 30-bp fragment is not shown). Heterozygous individuals will have both the 140-bp and the 110-bp fragments.

RFLP, PCR-RFLP, heteroduplex analysis, temporal temperature-gradient gel electrophoresis, single-strand conformation polymorphism (SSCP), SSP-PCR, Cleavase, bead array technology, and direct sequencing. The American College of Medical Genetics (ACMG) and the American College of Obstetricians and Gynecologists (ACOG) have recommended a core panel of 23 mutations that will identify 49% to 98% of carriers, depending on ethnic background. Sequencing panels include these and more rare variants.[10] Population differences and variable expressivity influence the choice of mutations to be covered.

Cytochrome P-450

Cytochrome P-450 comprises a group of enzymes localized to the endoplasmic reticulum (Fig. 12.13). These enzymes are mono-oxygenases; that is, they participate in enzymatic hydroxylation reactions and also transfer electrons to oxygen:

$$A{-}H + B{-}H_2 + O_2 \rightarrow A{-}OH + B + H_2O$$

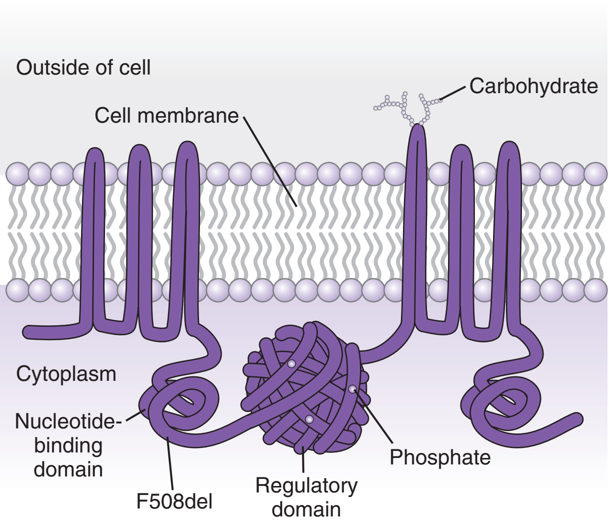

Outside of cell
Carbohydrate
Cell membrane
Cytoplasm
Nucleotide-binding domain
F508del
Regulatory domain
Phosphate

FIGURE 12.12 The CF transmembrane conductance regulator forms a channel in the cell membrane. The F508del and other mutations that affect its function are responsible for the phenotype of CF.

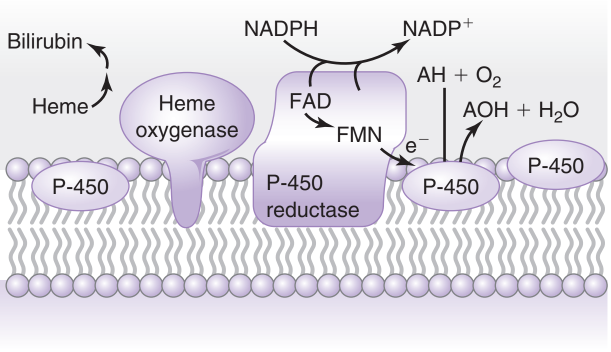

Bilirubin
Heme
Heme oxygenase
NADPH
NADP$^+$
FAD
FMN
e^-
AH + O_2
AOH + H_2O
P-450
P-450 reductase
P-450
P-450

FIGURE 12.13 NADPH cytochrome P-450 reductase catalyzes the reduction of NADPH and transfers electrons to flavin adenine dinucleotide (FAD). Electrons flow through FAD and flavin mononucleotide (FMN) to the cytochromes. The cytochromes then oxidize a variety of substrates (AH). NADPH cytochrome P-450 reductase also works with heme oxygenase to convert heme to biliverdin and eventually to bilirubin.

where A is the substrate and B is the hydrogen donor. These enzymes influence steroid, amino acid, and drug metabolism using NADH or NADPH as hydrogen donors. Oxygenation of lipophilic drugs renders them more easily excreted.

The cytochrome P-450 system is present in high concentrations in the liver and small intestine where the enzymes metabolize and detoxify compounds taken in orally (Fig. 12.14). The P-450 system varies from one person to another. This may in part account for different effects of drugs on different people. The metabolism of hormones, caffeine, chemotherapeutic drugs, antidepressants, and oral contraceptives is affected by these polymorphisms.[11,12] CYP-450 polymorphisms may also compound interactions of multiple drugs taken simultaneously.[13,14]

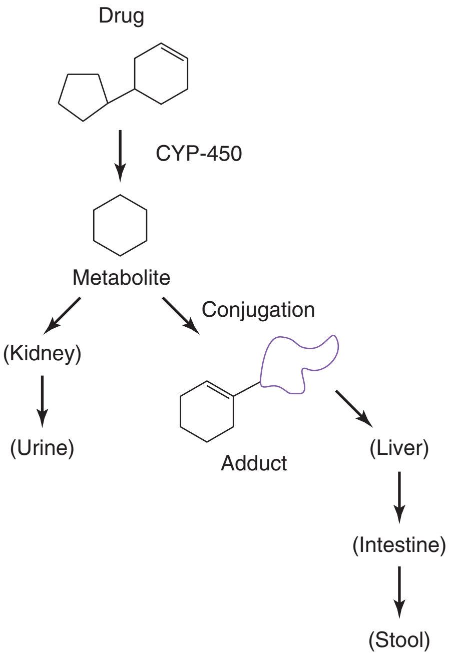


FIGURE 12.14 The oxidation activities of cytochrome P-450 proteins metabolize a variety of structurally diverse chemicals, including therapeutic drugs.

There are over 30 reported variations of CYP-450 enzymes.[15] Enzymes are classified according to families and subfamilies. For example, CYP2A6 is cytochrome P-450, subfamily IIA, polypeptide 6. CYP1A2 and the enzymes in the CYP2 and CYP3 families are considered to be most important in drug metabolism. Some of the enzymes reported to inhibit or induce drug metabolism include CYP1A2, CYP2A6, CYP2B6, CYP2C8, CYP2C9, CYP2C18, CYP2C19, CYP2D6, CYP2E1, CYP3A4, and CYP3A5-7. The genes coding for these enzymes are located throughout the genome.

Genetic polymorphisms of cytochrome P-450 genes are unequally distributed geographically and in different ethnic populations. Testing for these polymorphisms is used to predict the response to drugs sensitive to metabolism by this enzyme system. In the laboratory, testing for CYP-450 polymorphisms is performed by allele-specific PCR for particular polymorphisms. Multiple P-450 genetic variants may be screened by microarray, bead array, or sequencing.

SINGLE-GENE DISORDERS WITH NONCLASSICAL PATTERNS OF INHERITANCE

Mitochondrial mutations, genomic imprinting, and gonadal mosaicism do not follow Mendelian rules of inheritance. Mitochondrial mutations are inherited maternally (Fig. 12.15). Genomic imprinting is responsible for specific expression of genes in different cells and tissues. Imprinting is reset at meiosis and fertilization and is different in egg and sperm production.

Gonadal mosaicism is the generation of new mutations in germline cells. The mutated cells give rise to eggs or sperm carrying the mutation, which then

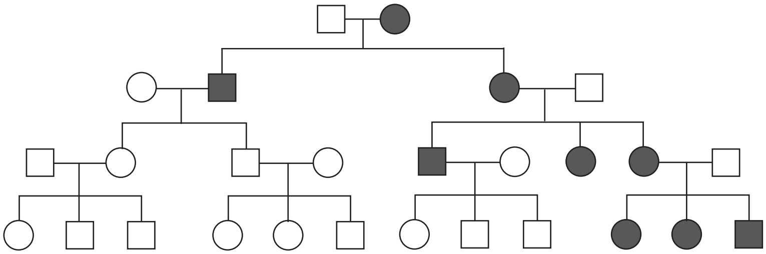


FIGURE 12.15 Mitochondrial mutations are maternally inherited.

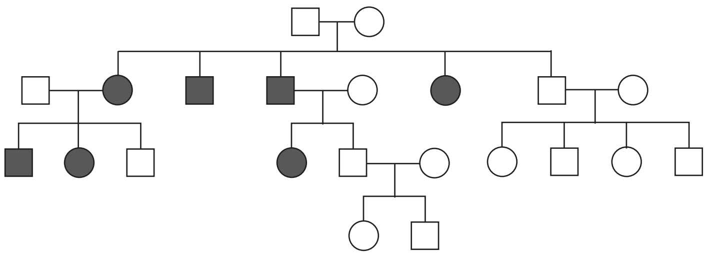


FIGURE 12.16 Gonadal mosaicism arises as a result of a new mutation. In this example, a dominant disease phenotype has been inherited from two unaffected parents. The mutation is present only in the germ cells of the first-generation parents but is inherited in all cells of the offspring.

becomes a heritable phenotype. Unusual pedigrees result (Fig. 12.16). Gonadal mosaicism is expected when phenotypically normal parents have more than one affected child (e.g., in osteogenesis imperfecta, an autosomal-dominant phenotype in a child from unaffected parents).

Advanced Concepts

An example of the nature of imprinting is illustrated by a comparison of mules and hinnies. A mule is the product of a male donkey and a female horse. A hinny is the product of a male horse and a female donkey. These animals are quite different in phenotype, even though they contain essentially the same genotype. Another illustration of the effects of imprinting is seen in animals cloned by nuclear transfer. Because this process bypasses the generation of eggs and sperm and fertilization, imprinting is not reset, and cloned animals display unexpected phenotypes, such as larger size or early onset of age-related diseases.

Mutations in Mitochondrial Genes

Mitochondria are cellular organelles responsible for energy production. Mitochondria contain their own genome, a circular DNA molecule 16,569 bp in length (Fig. 12.17). The mitochondrial genome contains 37 genes, including a 12S and 16S rRNA, 22 tRNAs, and 13 genes required for oxidative phosphorylation. In addition, the mtDNA contains a 1000-nt control region that encompasses transcription and replication regulatory elements. A database of mitochondrial genes and mutations is available at http://www.MITOMAP.org.[16]

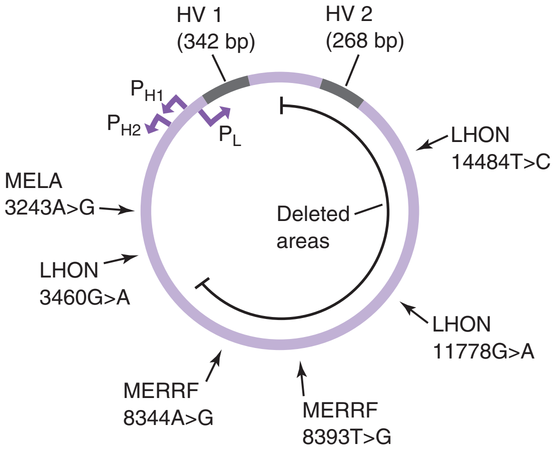


FIGURE 12.17 Mutations and deletions throughout the mitochondrial genome are associated with muscular and neurological disorders.

Mutations in mitochondrial genes affect energy production and are therefore manifested as diseases in the most energy-demanding organs, the muscles and the nervous system.[17] Mutations in several genes in the mitochondrial genome have been defined. These mutations result in a number of disease syndromes involving muscular and neurological disorders (Table 12.6).

Advanced Concepts

Disease severity depends on the number of mitochondria affected. Mitochondria are actively undergoing turnover in the cell, including fission and fusion, which are part of the mitochondrial network quality control. Alterations in mitochondrial dynamics can cause neuropathies or optic atrophy.[18]

Mitochondrial mutations are easily detected by a variety of molecular methods. Southern blot is used for detecting large deletions (Fig. 12.18). Point mutations are

TABLE 12.6 Diseases Resulting From Mutations in Mitochondrial Genes

Disorder	Gene Affected	Molecular Methodology
Kearns–Sayre syndrome	2–7 kb deletions	Southern blot, PCR, PCR-RFLP
Pearson syndrome	Deletions	Southern blot analysis of leukocytes, PCR-RFLP
Pigmentary retinopathy, chronic progressive external ophthalmoplegia	tRNA (tyr) deletion, deletions	PCR-RFLP, Southern blot analysis of muscle biopsy
Leber hereditary optic neuropathy	Cyt6 and URF* point mutations	PCR-RFLP
Mitochondrial myopathy, encephalopathy, lactic acidosis, and strokelike episodes	tRNA (leu) point mutations	PCR-RFLP, sequencing
Myoclonic epilepsy with ragged red fibers	tRNA (lys) point mutations	PCR-RFLP
Deafness		
Neuropathy, ataxia, retinitis pigmentosa (NARP)	ATPase VI point mutation	PCR-RFLP
Subacute necrotizing encephalomyelopathy with neurogenic muscle weakness, ataxia, retinitis pigmentosa (Leigh syndrome with NARP)	ATPase VI, NADH:ubiquinone oxidoreductase subunit mutations	PCR-RFLP
Mitochondrial DNA depletion syndrome	Thymidine kinase gene mutations	PCR, sequencing

*Unknown reading frame

analyzed by PCR-RFLP (Fig. 12.19). Interpretation of mutation analysis has long been complicated, however, by the extent of heteroplasmy (mutated mitochondria and normal mitochondria in the same cell) and the nature of the mutation.[19] A range of phenotypes may be present, even in the same family.

Genes that control mitochondrial functions are also found on the nuclear genome (Table 12.7). Unlike mitochondrial mutations that display maternal inheritance, these disorders have autosomal patterns of inheritance. Although the causative gene mutation is located on a nuclear gene, analysis of mitochondria may still show deletions or other mutations caused by the loss of the nuclear gene function.

Nucleotide-Repeat Expansion Disorders

Nucleotide repeats include short tandem repeats (STRs) with 1 to 10-bp repeating units. During DNA replication and meiosis, these STRs can expand (or contract) in length. Triplet-repeat mutations (expansions of STR with 3-bp repeating units) occur in coding and noncoding sequences of genes. The most well-known examples of triplet-repeat expansion diseases are fragile X syndrome and Huntington disease.

Advanced Concepts

In addition to family history, clinical tests including electroencephalography, neuroimaging, cardiac electrocardiography, echocardiography, magnetic resonance spectroscopy, and exercise testing are important for the accurate diagnosis of mitochondrial disorders. High blood or cerebrospinal fluid lactate concentrations, as well as high blood glucose levels, are observed in patients with some mitochondrial diseases. More direct tests, such as histological examination of muscle biopsies and respiratory chain complex studies on skeletal muscle and skin fibroblasts, are more specific for mitochondrial dysfunction.

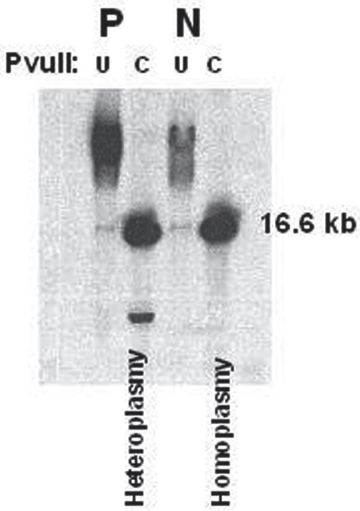


FIGURE 12.18 A mitochondrial deletion as revealed by Southern blot. DNA was cut with *Pvu*II, a restriction enzyme that cuts once in the mitochondrial genome. The membrane was probed for mitochondrial sequences. Normal mitochondria (N) yield one band at 16.6 kb when cut with *Pvu*II (C). Supercoiled, nicked, and a few linearized mitochondrial DNA circles can be seen in the uncut DNA (U). DNA from a patient with Kearns–Sayres syndrome (P) yields two mitochondrial populations, one of which has about 5 kb of the mitochondrial genome–deleted sequences. Because both normal and mutant mitochondria are present, this is a state of heteroplasmy. *(Photo courtesy of Dr. Elizabeth Berry-Kravis, Rush University Medical Center.)*

Fragile X Syndrome

Fragile X syndrome is associated with a triplet-repeat (CGG) expansion in the noncoding region 5′ to the fragile X mental retardation gene, *FMR-1* (Fig. 12.20). The expansion becomes so large in full fragile X syndrome (more than 2,000 CGG repeats) that the region is microscopically visible (Fig. 12.21). The CGG repeat expansion 5′ to the *FMR-1* gene also results in methylation of the region and transcriptional shutdown of

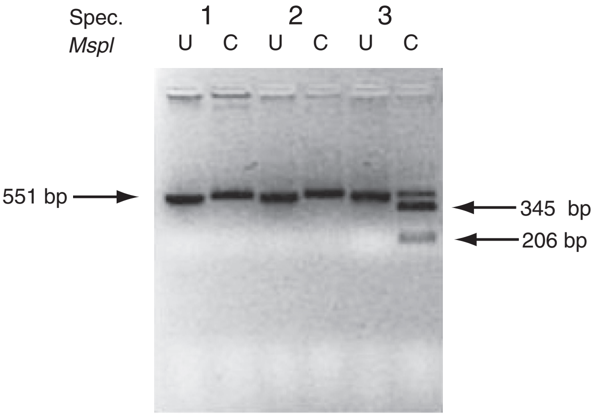


FIGURE 12.19 Detection of the NARP mitochondrial point mutation (ATPase VI 8993T>C or T>G) by PCR-RFLP. The PCR product was digested with the enzyme *Msp*I (C) or undigested (U). If the mutation is present, the enzyme will cut the PCR product into two pieces, as seen is specimen 3. *(Photo courtesy of Dr. Elizabeth Berry-Kravis, Rush University Medical Center.)*

TABLE 12.7 Some Disorders Caused by Nuclear Gene Mutations

Disorder	Gene Affected (Location)	Molecular Methodology
Mitochondrial DNA depletion syndrome	Succinate-CoA ligase, ADP-forming, beta subunit, SUCLA 2 (13q12.2-q13)	Southern blot
Mitochondrial neurogastrointestinal encephalomyopathy	Platelet-derived endothelial cell growth factor, ECGF (22q13-qter)	Sequencing
Progressive external ophthalmoplegia	Chromosome 10 open reading frame 2, C10ORF2 (10q24); polymerase, DNA, gamma, POLG (15q25); solute carrier family 25 (mitochondrial carrier), member 4, SLC25A4/ANT1 (4q25)	Southern blot, SSCP, sequencing

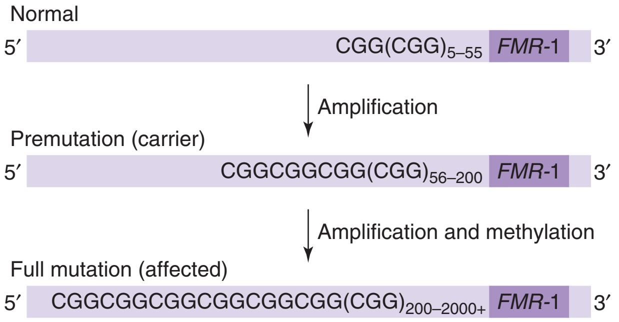


FIGURE 12.20 Triplet-repeat (CGG) expansions in sequences 5′ to the *FMR-1* gene are observed in fragile X carriers (up to 200 repeats) and fully affected individuals (more than 200 repeats). Normally there are fewer than 60 repeats. Expansion results from amplification of the triplet sequences during meiotic recombination events. The very large repeats (more than 200 repeats) are methylated on the C residues. This methylation turns off *FMR-1* transcription.

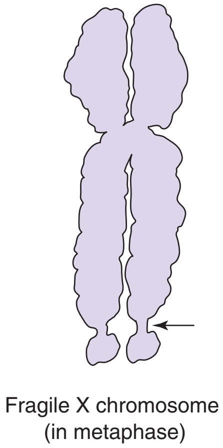


FIGURE 12.21 The fragile X chromosome is characterized by a threadlike process just at the telomere of the long arm (arrow). This is the site of disorganization of chromatin structure by the GC-rich repeat expansions.

FMR-1. CGG repeats can be interrupted by AGG, which dampens the methylation and silencing of *FMR-1*.

Symptoms of fragile X syndrome include learning disorders and mental retardation (IQ ~20), long face, large ears, and macroorchidism (large genitalia). Symptoms are more apparent at puberty. Symptoms increase in severity with each generation in a fragile X family (Fig. 12.22). Approximately 20% of women with *FMR-1* premutation (PM) will develop fragile X

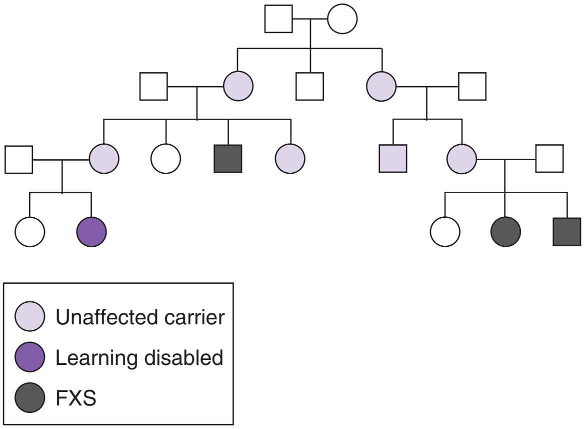


FIGURE 12.22 The symptoms of fragile X syndrome (FXS) become more severe with each generation. The fragile X chromosome cannot be transmitted from fathers to sons but can be transmitted from grandfathers to grandsons through their daughters.

primary ovarian insufficiency (FXPOI), with amenorrhea, menopausal follicle-stimulating hormone (FSH) levels, and possible estrogen deficiency.[20] Another condition associated with PM status is fragile X tremor and ataxia syndrome (FXTAS), with declining overall cognitive abilities with age.[21]

In addition to the fragile X chromosome observed by karyotyping, the state of the repeat expansion is also analyzed using PCR and by Southern blot (Fig. 12.23). **Premutations** in fragile X carriers are easily detected by PCR, with full mutations detectable by triple-repeat-primed PCR. AGG interruptions can be detected with AGG-triplet-repeat primers.[22] Southern blot can reveal cases of mosaicism where both premutations and full fragile X chromosomes are present in separate cell populations from the same patient.

Capillary electrophoresis is an increasingly popular option for identifying expanded FMR alleles, replacing the gel procedures (Fig. 12.24). Peak patterns indicate the presence of normal, premutation, and full fragile X mutation. The presence of AGG interrupters of the CGG repeats shows as gaps in the series of peaks. AGG promotes stability or slower expansion of the repeat region. The capillary electrophoresis method is faster and technically less demanding than Southern blot; however, the latter procedure may still be required to identify the presence of deletions within the *FMR-1* gene or the 5′ repeats in a percentage of the cells (mosaicism).

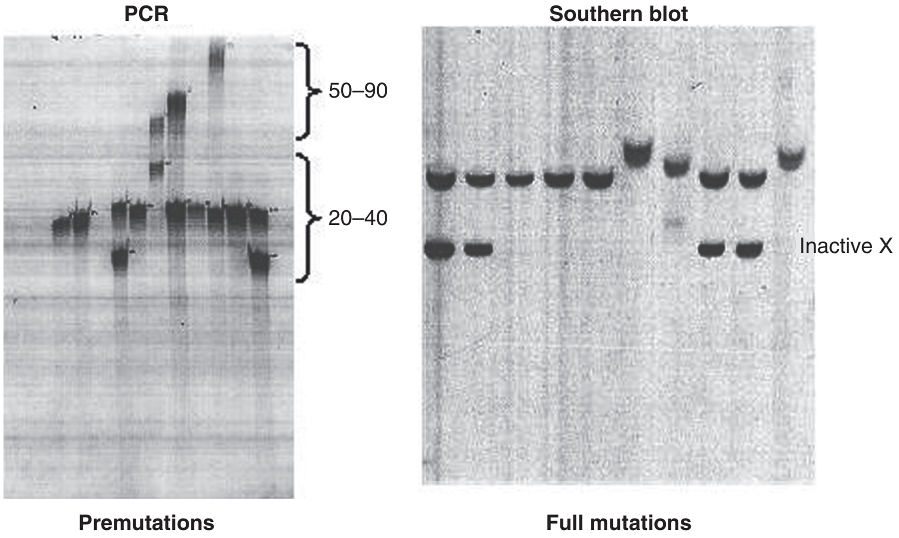


FIGURE 12.23 Detection of premutations by PCR *(left)* and full fragile X mutations by Southern blot *(right)*. Primers (one of which is labeled with ^{32}P) flanking the repeat region are used to generate labeled PCR products. Premutations appear as large amplicons in the 50- to 90-repeat range on the autoradiogram at left. Normal samples fall in the 20- to 40-repeat range. Full fragile X repeats are too large and GC rich to detect by standard PCR. Southern blots reveal full mutations in three of the samples shown. The inactive (methylated) X chromosome in four female patients is detected by cutting the DNA with a methylation-specific restriction enzyme. *(Photos courtesy of Dr. Elizabeth Berry-Kravis, Rush University Medical Center.)*

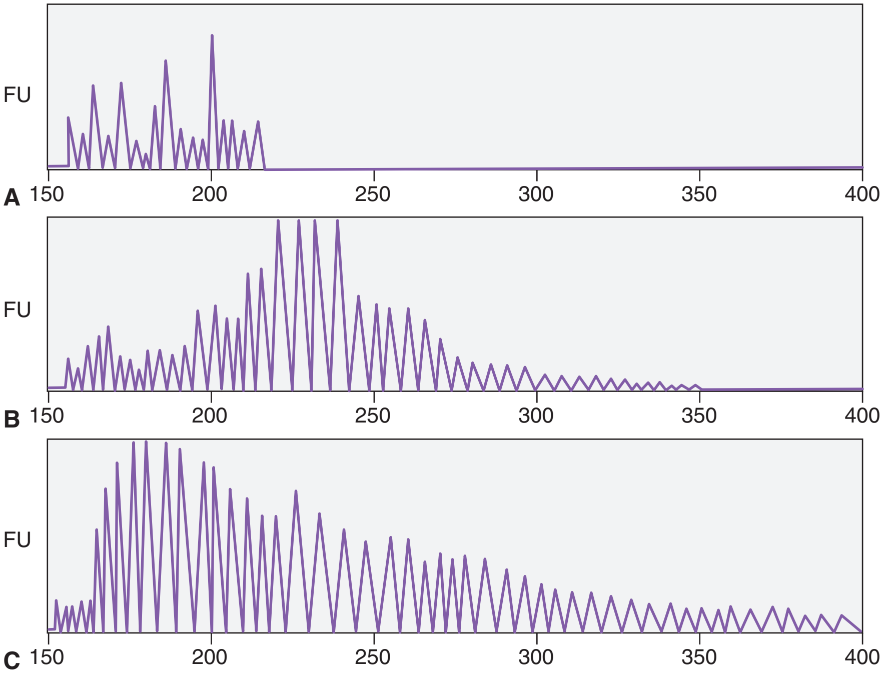


FIGURE 12.24 Fragile X premutations and full mutations appear as altered peak patterns in an electropherogram. PCR products of the *FMR-1* CGG repeat region in the normal X chromosome (A) are detected as peaks of less than 250 bases (A), whereas the premutation (B) and full mutation (C) are visible as regular peaks extending up to 400 bases.

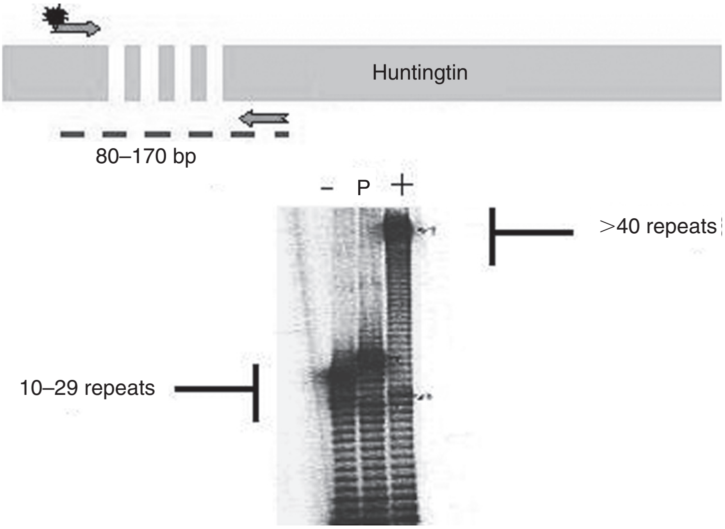


FIGURE 12.25 The huntingtin repeat expansion occurs within the coding region of the huntingtin gene. The expansion is detected directly by PCR using primers flanking the expanded region *(top)*. A ^{32}P-labeled primer is used, and the bands are detected by autoradiography of the polyacrylamide gel *(bottom)*. In this example, PCR products from the patient (P) fall within the normal range with the negative control (–). The positive control (+) displays the band sizes expected in Huntington disease. *(Photos courtesy of Dr. Elizabeth Berry-Kravis, Rush University Medical Center.)*

Huntington Disease

Huntington disease, first described by George Huntington in 1872, is associated with expansion within the huntingtin structural gene (4p16.3). In this repeat expansion, the sequence CAG expands from 9 to 37 repeats to 38 to 86 in the huntingtin gene. The clinical symptoms of Huntington disease include impaired judgment, slurred speech, difficulty in swallowing, abnormal body movements (chorea), personality changes, depression, mood swings, unsteady gait, and an intoxicated appearance. With onset in the 30s or 40s, these symptoms do not become obvious until the fourth or fifth decade of life, usually after family planning. The child of a person with Huntington disease has a 50% chance of inheriting the disorder. Genetic counseling, therefore, is important for younger persons with family histories of this disease, especially with regard to having children.

Advanced Concepts

The FMR protein (FMRP) binds RNA and is associated with polysomes. FMRP regulates translation of its bound mRNAs through alternative mRNA splicing, mRNA stability, and mRNA trafficking from the nucleus to the cytoplasm. FMRP may be associated with the miRNA pathway as well, preventing helicase-mediated miRNA suppression.[23,24]

In contrast to fragile X, where the repeat expansion takes place 5′ to the coding sequences, the Huntington expansion occurs within the coding region of the gene. The triplet expansion inserts multiple glutamine residues in the 5′ end of the huntingtin protein. This causes the protein to aggregate in plaques, especially in nervous tissue, causing the neurological symptoms seen in this disease. The expansion does not reach the size of the fragile X expansion and is detectable by standard PCR methods and capillary electrophoresis (Fig. 12.25).

Idiopathic Congenital Central Hypoventilation Syndrome

Idiopathic congenital central hypoventilation syndrome (CCHS) is a rare pediatric disorder characterized by inadequate breathing while asleep. More-affected children may also experience hypoventilation while awake. CCHS occurs in association with an intestinal disorder called Hirschsprung disease and symptoms of diffuse autonomic nervous system dysregulation/dysfunction. A number of gene mutations have been observed in CCHS, including a polyalanine expansion of the paired-like homeobox (*PHOX2b*) gene (4p12)13. The PHOX2b protein is a transcription factor containing a domain (homeobox) similar to other proteins that bind DNA.

In CCHS, a triplet-repeat expansion occurs inside of the *PHOX2b* gene, resulting in the insertion of multiple alanine residues into the protein. The severity of the disease increases with increasing numbers of repeats. The expansion is detected by PCR (Fig. 12.26).[25]

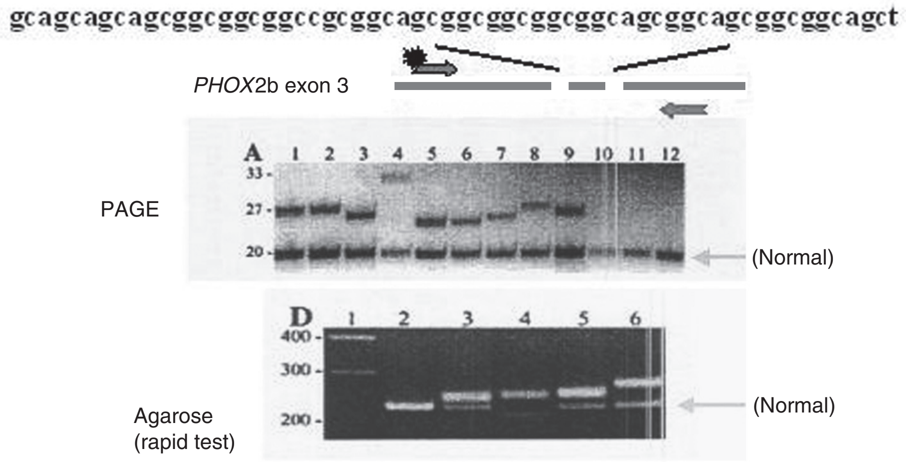


FIGURE 12.26 The triplet-repeat expansion of *PHOX2b* includes triplets that code for alanine *(top)*. The expansion is detected by PCR with a ^{32}P-labeled primer and polyacrylamide gel electrophoresis *(center)* or by standard PCR and agarose gel electrophoresis *(bottom)*. Normal specimens yield a single PCR product. CCHS specimens yield another larger product in addition to the normal product. The standard PCR test can rapidly show the presence of the expansion, and the PAGE test allows determination of the exact number of alanine codons that are present in the expansion. *(Photos courtesy of Dr. Elizabeth Berry-Kravis, Rush University Medical Center.)*

CCHS is usually apparent at birth. In some children, late onset of the disease occurs at 2 to 4 years of age. An estimated 62% to 98% of patients with CCHS have the *PHOX2b* gene mutation.[26]

Other Nucleotide Expansion Disorders

Fragile X, Huntington disease, and CCHS are three of a group of diseases caused by trinucleotide-repeat disorders. This category of diseases is subclassified into polyglutamine expansion disorders, which includes Huntington disease, and non–polyglutamine expansion disorders. Larger repeat units may also be involved in expansion mutations. Expansion of a hexanucleotide repeat unit is found in cases of amyotrophic lateral sclerosis (ALS).[27] Examples of repeat expansion diseases are listed in Table 12.8.

Genomic Imprinting

Genomic imprinting is transcriptional silencing through histone or DNA modification. Imprinting occurs during egg and sperm production and is different in DNA brought in by the egg or the sperm upon fertilization. The difference is exhibited in genetic disorders in which one or the other allele of a gene is lost.

A **uniparental disomy**/deletion demonstrates the nature of imprinting on chromosome 15. A deletion in the paternal chromosome 15, del(15)(q11q13), causes Prader–Willi syndrome. Symptoms of this disorder include mental disability, short stature, obesity, and hypogonadism. Loss of the same region from the maternal chromosome 15 results in Angelman syndrome, a disorder with very different symptoms, including ataxia, seizures, and inappropriate laughter. Both syndromes can occur in four ways: a deletion on the paternal or maternal chromosome 15, a mutation on the paternal or maternal chromosome 15, a translocation with loss of the critical region from one chromosome, and maternal or paternal uniparental disomy in which both chromosomes 15 are inherited from the mother and none from the father or vice versa.

Cytogenetic methods are used to test for these genetic lesions. Translocations and some deletions are detectable by standard karyotyping. High-resolution karyotyping can detect smaller deletions; however, other cases are not detectable microscopically. FISH with labeled probes to the deleted region can detect over 99% of

TABLE 12.8 Examples of Nonpolyglutamine Expansion Disorders

Disorder	Gene	Repeat	Expansion, (Normal)—(Symptomatic)*
Fragile XE	Fragile X mental retardation 2 (Xq28)	GCC	(6–35) – (over 200)
Friedreich ataxia	Frataxin, FRDA or X25 (9q13)	GAA	(7–34) – (over 100)
Myotonic dystrophy	Dystrophia myotonica protein kinase (9q13.2–13.3)	CTG	(5–37) – (over 50)
Spinocerebellar ataxia type 8	Spinocerebellar ataxia type 8 (13q21)	CTG	(16–37) – (110–250)
Spinocerebellar ataxia type 12†	Spinocerebellar ataxia type 12 (5q31–33)	CAG	(7–28) – (66–78)
Amyotrophic lateral sclerosis, frontotemporal disorder	C9 open reading frame 72 (9p21.1)	GGGGCC	(2–20) – (over 100)

*The phenotypic effects of intermediate numbers of repeats is not known.
†Although CAG codes for glutamine, this expansion occurs outside of the coding region of this gene and is not translated.

cases. PCR of RFLP or STR analysis has also been used to demonstrate uniparental disomy. Because imprinting (DNA methylation) is different on maternal and paternal chromosomes, methylation-specific PCR and Southern blot using methylation-specific restriction enzymes can aid in the diagnosis of these disorders. Assays developed for the detection of copy-number variants, including FISH, array-based comparative genomic hybridization (aCGH), and next-generation sequencing (NGS), have been used for the detection of genome-wide uniparental disomy (UPD) associated with constitutional and neoplastic disorders.[27]

Multifactorial Inheritance

Most disorders (and normal conditions) are controlled by multiple genetic and environmental factors. Multifactorial inheritance is displayed as a continuous variation in populations, with a normal distribution, rather than as a specific inheritance pattern. Nutritional or chemical exposures alter this distribution. The range may be discontinuous, with a threshold of manifestation. The phenotypic expression is conditioned by the number of controlling genes inherited. The chance of a first-degree relative having a similar phenotype is 2% to 7%.

High-resolution array methods and next-generation sequencing have further defined the genetic components of multifactorial illnesses. Large databases such as ClinVar and dbSNP aid in the interpretation of combinations of genetic (and somatic) variants. Further annotation of demographics, such as ethnicity or gender, and lifestyles, such as smoking or diet, will further add to the prognostic and diagnostic value of gene mutation analysis in these cases.

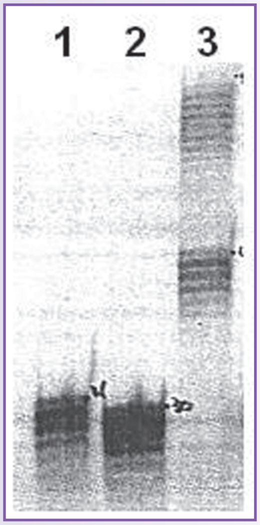
LIMITATIONS OF MOLECULAR TESTING

Although molecular testing for inherited diseases is extremely useful for early diagnosis and genetic counseling, there are circumstances in which genetic testing may not be the optimal methodology. To date, most therapeutic targets are phenotypic so that treatment is better directed to the phenotype. In genes with variable expressivity, finding a gene mutation may not predict the severity of the phenotype. For instance, clotting time and transferrin saturation are better guides for anticoagulant treatment than the demonstration of the causative gene mutations.

Molecular testing may discover genetic lesions in the absence of symptoms. This raises a possible problem as to whether treatment is indicated. This is increasingly possible with the use of large array or sequencing panels targeting hundreds of genes. Several genetic lesions may be present or polymorphic states of other normal genes may influence the disease state. Research methods and clinical trials using array technology and sequencing methods designed to scan at the genomic level may contribute to better diagnosis of complex diseases.

Case Study 12.1

A young mother was worried about her son. Having observed others, she was very aware of how her baby was expected to grow and acquire basic skills. As the child grew, however, he showed signs of slow development. His protruding ears and long face were becoming more noticeable as well. The pediatrician recommended chromosomal analysis for the mother and child. A constriction at the end of the X chromosome was found in the son's karyotype. The mother's karyotype was normal 46,XX. A Southern blot analysis for fragile X was performed on a blood specimen from the mother but showed no obvious abnormality. PCR analysis produced the following results:



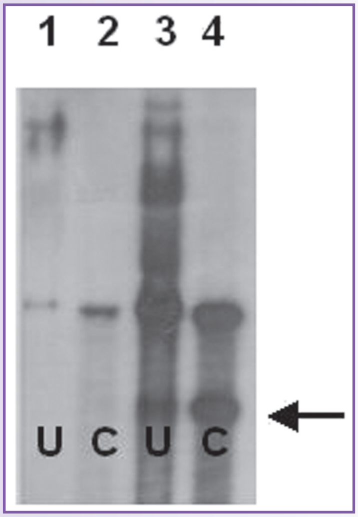
PCR analysis of the FMR promoter region showing two normal patterns (lanes 1 and 2) and the mother's pattern.

QUESTIONS:

1. *What do the PCR results in lane 3 indicate?*
2. *Why were there no abnormalities detected by Southern blot?*
3. *How would this result be depicted using triplet-primed PCR and capillary electrophoresis?*

Case Study 12.2

A 14-year-old girl with muscle weakness and vision difficulties (retinopathy) was referred for clinical tests. A muscle biopsy was performed, and aberrant mitochondria were observed in thin sections. Histochemical analysis of the muscle tissue revealed cytochrome oxidase deficiency in the muscle cells. A skeletal muscle biopsy specimen was sent to the molecular genetics laboratory for analysis of mitochondrial DNA. Southern blot analysis of *Pvu*II cut mitochondrial DNA exhibited a band at 16,000 bp in addition to the normal mitochondrial band at 16,500 bp, as follows:



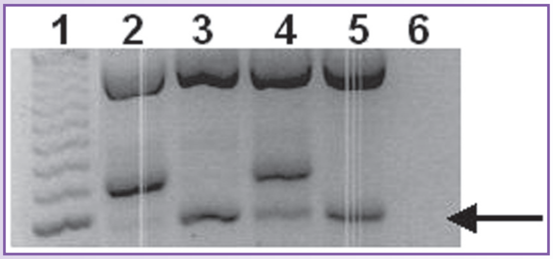
Southern blot of mitochondrial DNA uncut (U) and cut with *Pvu*II (C). Lanes 1 and 2, normal control; lanes 3 and 4, patient. Arrow points to a 16,000-kb band.

QUESTIONS:

1. *What is the 16,000 bp product? What is the 11,500 bp product?*
2. *What condition is associated with a 5 kb mitochondrial deletion?*
3. *Is this a case of homoplasmy or heteroplasmy?*

Case Study 12.3

A 40-year-old man was experiencing increasing joint pain, fatigue, and loss of appetite. He became alarmed when he suffered heart problems and consulted his physician. Routine blood tests revealed high serum iron (900 µg/dL) and 80% transferrin saturation. Total iron-binding capacity was low (100 µg/dL). The man denied any extensive alcohol intake. A blood specimen was sent to the molecular genetics laboratory for analysis. The following results were produced:



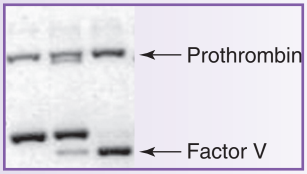
PCR-RFLP analysis of exon 4 of the *HFE* gene. The PCR product contains 1 recognition site for the restriction enzyme, *Rsa*I. An additional *Rsa*I site is created by the C282Y mutation, the most common inherited mutation in hemochromatosis. This extra site results in fragments of 240 bp, 110 bp (arrow), and ← 30 bp (not shown) instead of the 240-bp and 140-bp normal fragments. Lane 1, molecular-weight markers; lane 2, normal control; lane 3, homozygous C282Y; lane 4, heterozygous C282Y; lane 5, patient specimen; lane 6, reagent blank.

QUESTIONS:

1. *Does this patient have the C282Y mutation?*
2. *Is this mutation heterozygous or homozygous?*
3. *Based on these results, what is the likely explanation for the patient's symptoms?*

Case Study 12.4

A 30-year-old woman was brought to the emergency room with a painfully swollen left leg. She informed the nurses that she was taking contraceptives. Deep vein thrombosis was suspected, and compression ultrasound was ordered to look for pulmonary embolism. A clotting test for APC resistance resulted in a 1.5 ratio of clotting time with and without APC. An enzyme-linked immunosorbent assay (ELISA) test for D-dimer was positive. The patient was immediately treated with heparin, and blood samples were taken. A blood sample was sent to the molecular genetics laboratory to test for the factor V Leiden 1691 G→A mutation and prothrombin 20210 G→A mutations. The following results were produced:



Agarose gel electrophoresis showing the band pattern for prothrombin 20210/factor V Leiden mutation detection by multiplex SSP-PCR-RFLP. Specimens were cut with *Hind*III. Lane 1, molecular-weight marker; lane 2, normal control (prothrombin, 407 bp + 99 bp, factor V, 241 bp); lane 3, heterozygous control, prothrombin (407, 384, 99, 23 bp)/factor V Leiden (241, 209, 323 bp); lane 4, homozygous factor V Leiden, normal prothrombin; lane 5, homozygous prothrombin 20210 mutation/normal factor V; lane 6, patient specimen; lanes 7 and 8, normal specimens.

QUESTIONS:

1. *Does this patient have a clotting disorder?*
2. *Is there a factor V mutation? Is there a prothrombin mutation?*
3. *If present, is either mutation homozygous or heterozygous?*

STUDY QUESTIONS

1. Which of the following is not a triplet-repeat expansion disorder?
   - a. Fragile X syndrome
   - b. Huntington disease
   - c. Factor V Leiden
   - d. Congenital central hypoventilation syndrome

2. A gene was mapped to region 3, band 1, subband 1, of the long arm of chromosome 2. How would you express this location from an idiogram?

3. Which of the following can be detected by PCR?
   - a. Large mitochondrial deletions
   - b. Full fragile X disorder
   - c. Mitochondrial point mutations

4. A patient was tested for Huntington disease. PCR followed by PAGE revealed 25 CAG units. How should the results be interpreted?
   - a. This patient has Huntington disease.
   - b. This patient has a 1/25 chance of contracting Huntington disease.
   - c. This patient is normal at the Huntington locus.
   - d. The test is inconclusive.

5. Which of the following methods can detect the factor V Leiden mutation?
   - a. PCR-RFLP
   - b. SSP-PCR
   - c. Invader technology
   - d. All of the above

6. The most frequently occurring mutation in the *HFE* gene results in the replacement of cysteine (C) with tyrosine (Y) at position 282. How is this expressed according to the recommended nomenclature?

7. MELAS is a disease condition that results from an A to G mutation at position 3243 of the mitochondrial genome. This change creates a single *Apa*I restriction site in a PCR product, including the mutation site. What would you expect from a PCR-RFLP analysis for this mutation in a patient with MELAS?
   - a. A single PCR product resistant to digestion with *Apa*I
   - b. A single PCR product that cuts into two fragments upon digestion with *Apa*I
   - c. A single PCR product only if the mutation is present
   - d. Two PCR products

8. A father affected with a single-gene disorder and an unaffected mother have four children (three boys and a girl), two of whom (one boy and the girl) are affected. Draw the pedigree diagram for this family.

   D16S539, an STR, was analyzed in the family. The result showed that the father had the 6,8 alleles, and the mother had the 5,7 alleles. The affected children had the 5,6 and 6,7 alleles, and the unaffected children had the 5,8 and 7,8 alleles.
   - a. If D16S539 is located on chromosome 16, where is the gene for this disorder likely to be located?
   - b. To which allele of D16S539 is the gene linked?

   How might one perform DNA analysis for the presence of the disorder?
   - a. Analyze D16S539 for the 6 allele by PCR.
   - b. Sequence the entire region of the chromosome where D16S539 was located.
   - c. Test as many STRs as possible by PCR.
   - d. Use a variant-specific test to detect the unknown gene mutation.

9. Exon 4 of the *HFE* gene from a patient suspected of having hereditary hemochromatosis was amplified by PCR. The G to A mutation, frequently found in hemochromatosis, creates an *Rsa*1 site in exon 4. When the PCR products are digested with *Rsa*1, which of the following results would you expect to see if the patient has the mutation?
   - a. None of the PCR products will be cut by *Rsa*1.
   - b. There will be no PCR product amplified from the patient DNA.

- **c.** The patient's PCR product will yield extra bands upon *Rsa*1 digestion.
- **d.** The normal control PCR products will yield extra *Rsa*1 bands compared with the patient sample.

10. Most people with the C282Y or H63D *HFE* gene mutations develop hemochromatosis symptoms. This is a result of

- **a.** iron loss.
- **b.** excessive drinking.
- **c.** high penetrance.
- **d.** healthy lifestyle.
- **e.** glycogen accumulation.

11. The majority of disease-associated mutations in the human population are

- **a.** autosomal dominant.
- **b.** autosomal recessive.
- **c.** X-linked.
- **d.** found on the Y chromosome.

12. Bead array technology is most appropriate for which of the following?

- **a.** Cystic fibrosis mutation detection
- **b.** Chromosomal translocation detection
- **c.** STR linkage analysis
- **d.** Restriction fragment length polymorphisms

References

1. Anderson J. The Lyon hypothesis. *British Medical Journal* 1963;5367:1215–1216.
2. Davidson R. The Lyon hypothesis. *Journal of Pediatrics* 1964;65:765–775.
3. Goetzman E, Gong Z, Schiff M, Wang Y, Muzumdar RH. Metabolic pathways at the crossroads of diabetes and inborn errors. *Journal of Inherited Metabolic Disorders* 2017;41(1):5–17.
4. Huber S, McMaster KJ, Voelkerding KV. Analytical evaluation of primer engineered multiplex polymerase chain reaction–restriction fragment length polymorphism for detection of factor V Leiden and prothrombin G20210A. *Journal of Molecular Diagnostics* 2000;2:153–157.
5. Agarwal R, Peters SM, Shemirani M, von Ahsen N. Improved real-time multiplex polymerase chain reaction detection of methylenetetrahydrofolate reductase (MTHFR) 677c>t and 1298a>c polymorphisms using nearest neighbor model-based probe design. *Journal of Molecular Diagnostics* 2007;9:345–350.
6. Liew M, Wittwer C, Voelkerding KV. Nucleotide extension genotyping by high-resolution melting. *Journal of Molecular Diagnostics* 2010;12:731–738.
7. Kerem B, Buchanan JA, Durie P, Corey ML, Levison H, Rommens JM, Buchwald M, Tsui L-C. DNA marker haplotype association with pancreatic sufficiency in cystic fibrosis. *American Journal of Human Genetics* 1989;44:827–834.
8. De Boecka K, Zolin H, Cuppens H, Olesen HV, Viviani L. The relative frequency of CFTR mutation classes in European patients with cystic fibrosis. *Journal of Cystic Fibrosis* 2014;13:403–409.
9. McGinniss M, Chen C, Redman JB, Buller A, Quan F, Peng M, Giusti R, Hantash FM, Huang D, Sun W, Strom CM. Extensive sequencing of the *CFTR* gene: lessons learned from the first 157 patient samples. *Human Genetics* 2005;118:331–338.
10. Elliott A, Radecki J, Moghis B, Li X, Kammesheidt A. Rapid detection of the ACMG/ACOG-recommended 23 CFTR disease-causing mutations using ion torrent semiconductor sequencing. *Journal of Biomolecular Technology* 2012;23:24–30.
11. Haraya K, Kato M, Chiba K, Sugiyama Y. Prediction of inter-individual variability on the pharmacokinetics of CYP1A2 substrates in non-smoking healthy volunteers. *Drug Metabolism and Pharmacokinetics* 2016;31:276–284.
12. Parra-Guillen Z, Berger PB, Haschke M, Donzelli M, Winogradova D, Pfister B, Früh M, Gillessen S, Krähenbühl S, Kloft C, Joerger M. Role of Cytochrome P450 3A4 and 1A2 phenotyping in patients with advanced non-small-cell lung cancer receiving Erlotinib treatment. *Basic & Clinical Pharmacology & Toxicology* 2017;121:309–315.
13. McDonald M, Au NT, Rettie AE. P450-based drug-drug interactions of amiodarone and its metabolites: diversity of inhibitory mechanisms. *Drug Metabolism and Disposition; The Biological Fate of Chemicals* 2015;43:1661–1669.
14. Granfors M, Backman JT, Laitila J, Neuvonen PJ. Oral contraceptives containing ethinyl estradiol and gestodene markedly increase plasma concentrations and effects of tizanidine by inhibiting cytochrome P450 1A2. *Clinical Pharmacology and Therapeutics* 2005;78:400–411.
15. Sim S, Ingelman-Sundberg M. The human cytochrome P450 (CYP) allele nomenclature website: a peer-reviewed database of CYP variants and their associated effects. *Human Genomics* 2010;4:278–281.
16. Lott M, Leipzig JN, Derbeneva O, Xie HM, Chalkia D, Sarmady M, Procaccio V, Wallace DC. mtDNA variation and analysis using mitomap and mitomaster. *Current Protocols in Bioinformatics* 2013;44:1.23–1.26.
17. Suomalainen A, Battersby BJ. Mitochondrial diseases: the contribution of organelle stress responses to pathology. *Nature Reviews Molecular and Cell Biology* 2017;19(2):77–92.
18. Suárez-Rivero J, Villanueva-Paz M, de la Cruz-Ojeda P, de la Mata M, Cotán D, Oropesa-Ávila M, de Lavera I, Álvarez-Córdoba M, Luzón-Hidalgo R, Sánchez-Alcázar JA. Mitochondrial dynamics in mitochondrial diseases. *Diseases* 2017;5:ii.
19. Moraes C, Atencio DP, Oca-Cossio J, Diaz F. Techniques and pitfalls in the detection of pathogenic mitochondrial DNA mutations. *Journal of Molecular Diagnostics* 2003;5:197–208.
20. Man L, Lekovich J, Rosenwaks Z, Gerhardt J. Fragile X-associated diminished ovarian reserve and primary ovarian

insufficiency from molecular mechanisms to clinical manifestations. *Frontiers in Molecular Neuroscience* 2017;10:290.

21. Wheeler A RM, Hagerman R, Mailick M, Riley C. Implications of the FMR1 premutation for children, adolescents, adults, and their families. *Pediatrics* 2017;139:S172–S182.

22. Hayward B, Usdin K Improved assays for AGG interruptions in fragile X premutation carriers. *Journal of Molecular Diagnostics* 2017;19:828–835.

23. Kenny P, Zhou H, Kim M, Skariah G, Khetani RS, Drnevich J, Arcila ML, Kosik KS, Ceman S. MOV10 and FMRP regulate AGO2 association with microRNA recognition elements. *Cell Reports* 2014;9:1729–1741.

24. Ascano MJ, Mukherjee N, Bandaru P, Miller JB, Nusbaum JD, Corcoran DL, Langlois C, Munschauer M, Dewell S, Hafner M, Williams Z, Ohler U, Tuschl T. FMRP targets distinct mRNA sequence elements to regulate protein expression. *Nature* 2012; 492:382–386.

25. Weese-Mayer D, Berry-Kravis EM, Zhou L, Maher BS, Silvestri JM, Curran ME, Marazita ML. Idiopathic congenital central hypoventilation syndrome: analysis of genes pertinent to early autonomic nervous system embryologic development and identification of mutations in *PHOX2b*. *American Journal of Medical Genetics* 2003;123A:267–278.

26. D'Alessandro V, Mason II TBA, Pallone MN, Patano J, Marcus CL. Late-onset hypoventilation without *PHOX2b* mutation or hypothalamic abnormalities. *Journal of Clinical Sleep Medicine* 2005;1(2):169–172.

27. Van Mossevelde S, Zee, Cruts M, VanBroeckhoven C. Relationship between C9orf72 repeat size and clinical phenotype. *Current Opinion in Genetics & Development* 2017;44:117–124.

Chapter 13

Molecular Oncology

Outline

CLASSIFICATION OF NEOPLASMS
MOLECULAR BASIS OF CANCER
ANALYTICAL TARGETS OF MOLECULAR TESTING
GENE AND CHROMOSOMAL MUTATIONS IN SOLID TUMORS
  Human Epidermal Growth Factor Receptor 2, HER2/neu/erb-b2 1 *(17q21.1)*
  Epidermal Growth Factor Receptor, EGFR *(7p12)*
  Kirsten Rat Sarcoma Viral Oncogene Homolog, K-ras *(12p12); Neuroblastoma ras,* N-ras *(1p13); and Harvey Rat Sarcoma Viral Oncogene Homolog,* H-ras *(11p15)*
  Ewing Sarcoma, EWS *(22q12)*
  Synovial Sarcoma Translocation, Chromosome 18—Synovial Sarcoma Breakpoint 1 and 2, SYT-SSX1, SYT-SSX2 *t(X;18)(p11.2;q11.2)*
  Paired Box-Forkhead in Rhabdomyosarcoma, PAX3-FKHR, PAX7-FKHR, *t(1;13), t(2;13)*
  Tumor Protein 53, TP53 *(17p13)*
  Ataxia Telangiectasia Mutated Gene, ATM *(11q22)*
  Breast Cancer 1 Gene, BRCA1 *(17q21), and Breast Cancer 2 Gene,* BRCA2 *(13q12)*
  Von Hippel–Lindau Gene, VHL *(3p26)*
  V-myc Avian Myelocytomatosis Viral-Related Oncogene, Neuroblastoma-Derived, MYCN *or* n-myc *(2p24)*
  V-Ros Avian UR2 Sarcoma Virus Oncogene Homolog 1 (ROS1) *Proto-Oncogene (6q22.1) and Rearranged During Transfection* (RET) *Proto-Oncogene (10q11)*
  Anaplastic Lymphoma Receptor Tyrosine Kinase (ALK) *Proto-Oncogene, 2p23.1*
  V-Kit Hardy-Zuckerman 4 Feline Sarcoma Viral Oncogene Homolog, KIT, c-KIT *(4q12)*
  Other Molecular Abnormalities
  Microsatellite Instability
  Loss of Heterozygosity
  Liquid Biopsy
MOLECULAR ANALYSIS OF LEUKEMIA AND LYMPHOMA
  Gene Rearrangements
    V(D)J Recombination
    Immunoglobulin Heavy-Chain Gene Rearrangement in B Cells
    Immunoglobulin Light-Chain Gene Rearrangement in B Cells
    T-Cell Receptor Gene Rearrangement
    Detection of Clonality
    Molecular Analysis of Immunoglobulin Heavy-Chain Gene Clonality
    Immunoglobulin Light-Chain Gene Rearrangements
    T-Cell Receptor Gene Rearrangements
    Banding Patterns
  Mutations in Hematological Malignancies
    Lymphoid Malignancies
    Myeloid Malignancies
  Mutation Spectra

Objectives

13.1 Present basic cancer biology and terminology.
13.2 Identify checkpoints in the cell division cycle that are critical for regulated cell proliferation.
13.3 List molecular targets that are useful for diagnosing and monitoring solid tumors.
13.4 Explain how microsatellite instability is detected.
13.5 Describe loss of heterozygosity and its detection.
13.6 Contrast cell-specific and tumor-specific molecular targets.
13.7 Show how clonality is detected using antibody and T-cell receptor gene rearrangements.
13.8 Describe translocations associated with hematological malignancies that can be used for molecular testing.
13.9 Interpret data obtained from the molecular analysis of patients' cells, and determine if a tumor population is present.

Oncology is the study of tumors. A tumor, or **neoplasm,** is a growth of tissue that exceeds that of normal tissue and is not coordinated with it. Tumors are either benign (not recurrent) or malignant (invasive and tending to recur at multiple sites). ***Cancer*** is a term that includes all malignant tumors. Molecular oncology is the study of cancer at the molecular level, using techniques that allow the direct detection of genetic alterations, down to single-base-pair changes.

CLASSIFICATION OF NEOPLASMS

Cancer is generally divided into two broad groups, solid tumors and hematological malignancies. Solid tumors are designated according to the tissue of origin as **carcinomas** (epithelial) or **sarcomas** (bone, cartilage, muscle, blood vessels, fat). **Teratocarcinomas** consist of multiple cell types. Benign tumors are named by adding the suffix -*oma* to the tissue of origin. For example, an adenoma is a benign glandular growth. An adenocarcinoma is malignant.

Metastasis is the movement of dislodged tumor cells from the original (primary) site to other locations. Only malignant tumors are metastatic. No one characteristic of the primary tumor predicts the likelihood of metastasis. Both tumor and normal cell factors are involved. The presence of metastasis increases the difficulty of treatment. Clinical analysis may be performed to detect the presence of relocated or circulating metastasized cells to aid in treatment strategy.

With regard to hematological malignancies, tumors arising from white blood cells are referred to as leukemias and lymphomas. **Leukemia** is a neoplastic disease of blood-forming tissue in which large numbers of white blood cells populate the bone marrow and peripheral blood. **Lymphoma** is a neoplasm of lymphocytes that forms discrete tissue masses. The difference between these diseases is not clear because lymphocytic leukemias and lymphomas can display bone marrow and blood symptoms similar to those of leukemias. Furthermore, chronic lymphomas can progress to leukemia. Conversely, leukemias can display lymphomatous masses without overpopulation of cells in the bone marrow.

Within lymphomas, Hodgkin (or Hodgkin's) disease is histologically and clinically different from all other types of lymphoma, termed non-Hodgkin (non-Hodgkin's) lymphoma (NHL). **Plasma cell** neoplasms, which arise from terminally differentiated B cells, are also classified in a separate category. Some of the physiological symptoms of plasma cell tumors are related to the secretion of immunoglobulin fragments by these tumors.

Leukemias arising from undifferentiated cells in the bone marrow are classified as myeloid diseases, such as acute and chronic myeloid leukemia. Myelodysplastic syndrome is a dysregulation of cells in a variety of differentiation states.

Advanced Concepts

As imaging technology advanced, several efforts were made in the classification of NHL. The earliest was the Rappaport classification in 1966, developed at the Armed Forces Institute of Pathology. The Keil classification, used in Europe, and the Lukes and Collins classification, used in the United States, were proposed in 1974. In 1982 an international group of hematopathologists proposed the Working Formulation for Clinical

Usage for classification of NHL.[1] The Working Formulation was revised in 1994 by the International Lymphoma Study Group, which proposed the World Health Organization/Revised European-American Classification of Lymphoid Neoplasms (REAL). The REAL classification includes genetic characteristics in addition to morphological tissue architecture. With increasing ability to detect molecular characteristics of cells, including patterns of gene expression, classification will continue to evolve.[2,3]

MOLECULAR BASIS OF CANCER

Cancer is caused by nonlethal mutations in DNA. The mutations affect two types of genes: **oncogenes** and **tumor-suppressor** genes. These genes control the **cell division cycle** and cell survival (Fig. 13.1).

Oncogenes promote cell division. Oncogenes include cell membrane receptors that are bound by growth factors, hormones, and other extracellular signals. These receptors transduce signals through the cell membrane into the cytoplasm through a series of protein modifications that ultimately reach the nucleus and activate factors that initiate DNA synthesis (moving the cell from G1 to S phase of the cell cycle) or mitosis (moving from G2 to M). Oncogenes also support cell survival by inhibiting **apoptosis,** or self-directed cell death. More than 100 oncogenes have been identified in the human genome.

Tumor suppressors include factors that control transcription, or the translation of genes required for cell division. They also participate in repairing DNA damage and in promoting apoptosis. Tumor suppressors slow down or stop cell division by counteracting the movement of the cell from G1 to S or G2 to M phase. These two points are therefore referred to as the G1 **checkpoint** and G2 checkpoint in the cell division cycle. More than 30 tumor-suppressor genes have been identified.

In cancer cells, mutations in oncogenes are usually gain-of-function mutations, resulting from amplification or translocation of DNA regions containing the genes or activating mutations that cause aberrant activity of the proteins. Mutations in tumor-suppressor genes are usually loss-of-function mutations, resulting

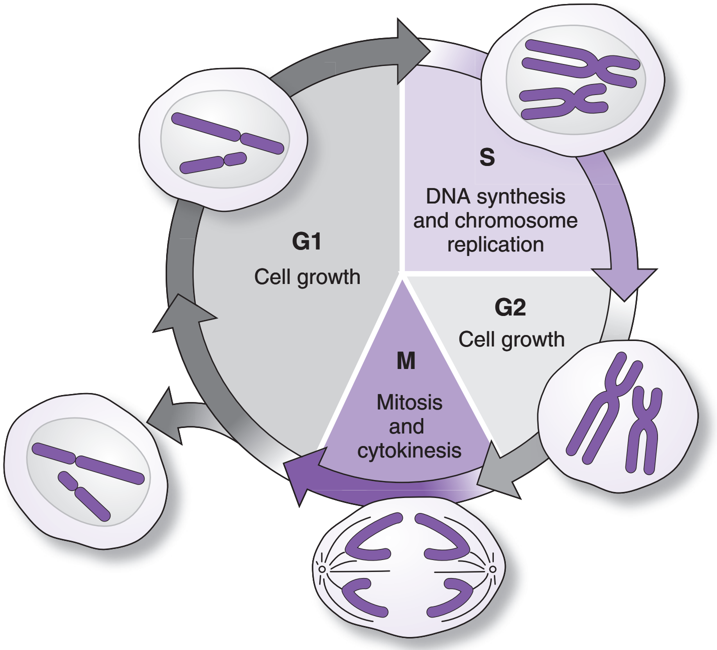


FIGURE 13.1 The cell division cycle. After mitosis (M), there are two haploid (one diploid) complements of chromosomes (46 chromosomes) in the G1 phase of the cell division cycle. DNA is replicated during the S phase, resulting in four haploid (two diploid) complements in the G2 phase. The chromosomes are distributed to two daughter cells at mitosis, with each receiving 46 chromosomes. Cancer results when the cell division cycle proceeds from G1 to S or G2 to M phase inappropriately.

in inactivation of the tumor-suppressor gene products. These mutations may occur through deletion, translocation, or mutation of the genes. Molecular laboratory testing aids in the diagnosis and treatment strategies for tumors by detecting abnormalities in specific tumor suppressors or oncogenes.

ANALYTICAL TARGETS OF MOLECULAR TESTING

Although not yet completely standardized, several tests are performed in almost every molecular pathology laboratory. These tests assess tissue-specific and tumor-specific targets. Tissue-specific targets are molecular characteristics of the tissue from which a tumor arose. The presence of DNA, RNA, protein, or other molecules from these targets in abnormal amounts or locations is used to detect and monitor the presence of the tumor. For example, molecular tests are designed to detect DNA or RNA from cytokeratin genes in gastric cancer, carcinoembryonic antigen in breast cancer, and rearranged immunoglobulin or T-cell receptor genes in lymphoma. Although tissue-specific markers are useful, they are also expressed by normal cells, and their presence does not always prove the presence of cancer.

In contrast, tumor-specific targets are not present in normal cells and are, therefore, more definitive with respect to the detection of a tumor. Tumor-specific genetic structures result from genome, chromosomal, or gene abnormalities in oncogenes and tumor-suppressor genes that are associated with the development of the tumor. Gene mutations and chromosomal translocations are found in solid tumors, leukemias, and lymphomas. Cell free nucleic acid or circulating tumor cells carrying oncogenic mutations can be detected in blood and other body fluids. Genome mutations, or **aneuploidy,** result in part from the loss of coordinated DNA synthesis and cell division that occurs when tumor suppressors or oncogenes are dysfunctional. The following sections describe procedures commonly performed in molecular pathology laboratories; however, due to the rapid advances in this area, the descriptions cannot be all-inclusive. The discussion is divided into solid tumor testing and testing for hematological malignancies. As will be apparent, however, some tests are applicable to both types of malignancies.

GENE AND CHROMOSOMAL MUTATIONS IN SOLID TUMORS

Molecular tests are routinely performed to aid in the diagnosis, characterization, and monitoring of solid tumors. Some of these tests have been part of molecular pathology for many years. Others are relatively new to the clinical laboratory. The methods applied to detect the molecular characteristics of tumors are described in previous chapters.

Human Epidermal Growth Factor Receptor 2, *HER2/neu/erb-b2 1* (17q21.1)

HER2/neu was discovered in rat neuro-/glioblastoma cell lines in 1985.[1] Later it was found to be the same gene as the avian erythroblastic leukemia viral oncogene homolog 2, or *ERBB2*. Its gene product is a 185-kd cell membrane protein that adds phosphate groups to tyrosines on itself and other proteins (tyrosine-kinase activity). This receptor is one of several transmembrane proteins with tyrosine-kinase activity (Fig. 13.2). It is very similar to a family of epidermal-growth-factor receptors that is overexpressed in some cancers[2] (Fig. 13.3). In normal cells, this protein is required for cells to grow and divide. *HER2/neu* is overexpressed in 25% to 30% of human breast cancers, in which overexpression of *HER2/neu* is a predictor of a more aggressive growth and metastasis of the tumor cells. It is also an indication for the use of anti-*HER2/neu* antibody drug therapy, which works best on tumors overexpressing *HER2/neu*. Herceptin therapy is indicated presently for women with *HER2/neu*-positive (*HER2/neu*-overexpressed) breast cancer.

Overexpression of the *HER2/neu* oncogene is performed by immunohistochemistry (IHC) using monoclonal and polyclonal antibodies to detect the HER2/neu protein. The IHC test works best on fresh or frozen tissue. Fluorescent in situ hybridization (FISH), which measures DNA and RNA of *HER2/neu,* is more reliable than IHC, especially in older, fixed tissue,[4] but is less readily available.

Southern, northern, and western studies have shown that *HER2/neu* gene amplification is highly correlated with the presence of increased *HER2/neu* RNA and protein.[5] In contrast to IHC, which measures increased amounts of HER2/neu protein, FISH directly detects

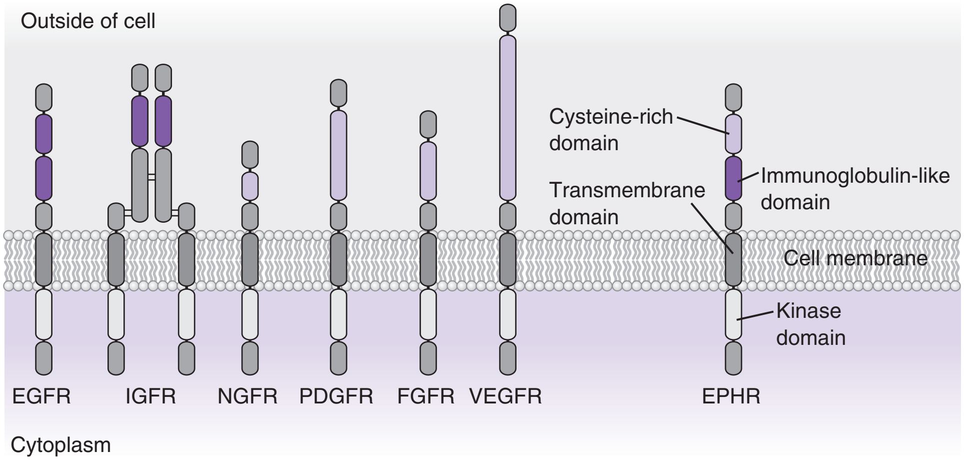


FIGURE 13.2 Receptor tyrosine kinases include epidermal growth factor receptor (EGFR), insulin growth factor receptor (IGFR), nerve growth factor receptor (NGFR), platelet-derived growth factor receptor (PDGFR), fibroblast growth factor receptor (FGFR), vascular endothelial growth factor receptor (VEGFR), and ephrin receptor (EPHR). These molecules share similarities in that they include a kinase domain, transmembrane domain, cysteine-rich domain, and immunoglobulin-like domain. The EPH receptor has two fibronectin type III domains.

increased copy numbers of the *HER2/neu* gene in DNA likely responsible for the increased protein. *HER2/neu* gene amplification occurs as a result of tandem duplication of the gene or other genetic events as the tumor cells continue to divide. FISH testing for *HER2/neu* gene amplification requires a labeled probe for the *HER2/neu* gene and a differently labeled control probe for the centromere of chromosome 17. For instance, a probe that spans the entire *HER2/neu* gene labeled in orange and a probe that binds to the centromere of chromosome 17 labeled in green should yield two green signals, each associated with an orange signal per nucleus. The copy number of *HER2/neu* relative to centromere 17 indicates whether *HER2/neu* is amplified (present in multiple copies on the same chromosome). Data are reported as a ratio of the number of HER2 signals to chromosome 17 centromere signals. A ratio of more than 2 is considered positive or amplified. The number of signals is enumerated in 50 to 100 cells.

Chromogenic in situ hybridization (CISH) is another method that has been used to detect *HER2/neu* gene amplification.[5,6] Using a standard bright-field microscope, CISH technology also detects deletions, translocations, or a change in the number of chromosomes. An attractive feature of CISH is that the slide images are permanent, facilitating documentation and consultations. Another bright-field imaging method, **silver-enhanced in situ hybridization (SISH),** introduced for the determination of *HER2* status, has a high correlation with FISH.[7] Initially, in situ hybridization with chromogenic detection by gene-specific probes was limited by high background and low signal intensity. Probes with repetitive sequences removed were designed to increase specificity. Although FISH, SISH, and CISH are more accurate and less subjective methods than IHC, IHC is faster, is less expensive, and allows the pathologist to assess target gene expression along with other visible landmarks on the slide. Furthermore, protein overexpression (detectable by IHC) can occur without gene amplification (detectable by ISH). Some laboratories use IHC as an initial screening method and then confirm the results with ISH.

Epidermal Growth Factor Receptor, *EGFR* (7p12)

Like *HER2/neu*, the epidermal growth factor receptor gene *(EGFR, ERBB1)* is a member of the *ERBB* family of growth factor receptors that also includes *ERB3/HER3* and *ERB4/HER4* (see Fig. 13.3). All of these proteins

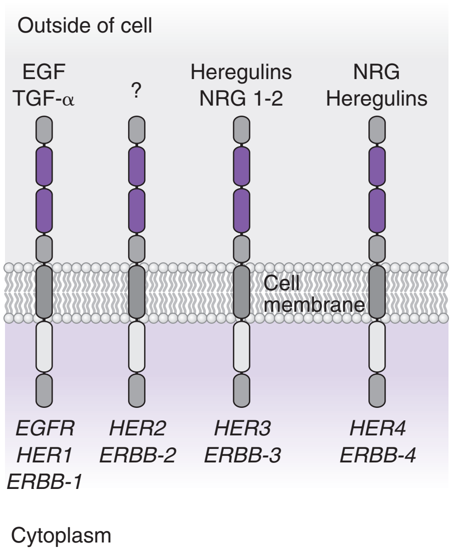


FIGURE 13.3 The ERBB family of growth factor receptors includes the HER2 receptor and EGFR. The factors that bind to these receptors on the cell surface begin a cascade of events, including autophosphorylation and phosphorylation of other proteins by the receptors. These factors include the epidermal growth factor (EGF), human transforming growth factor alpha (TGF-α), heregulins, and neuregulins (NRGs). EGF, NRGs, and heregulins are small peptides that are active in the development of various cell types, such as gastric mucosa, the heart, and the nervous system.

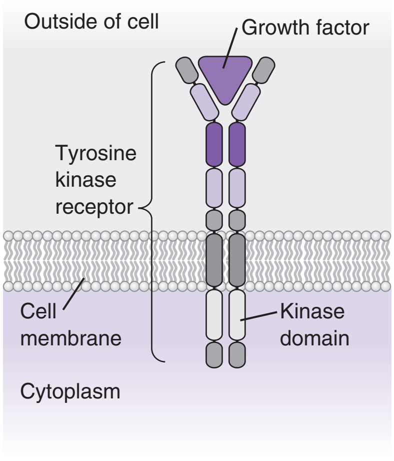


FIGURE 13.4 Upon binding by epidermal growth factor (EGF), the EGFR receptor in the cell membrane forms a dimer with itself or with other members of the ERBB family of receptors. The dimerization initiates a cascade of events, starting with phosphorylation of the receptor itself, catalyzed by the kinase domain.

are located in the cell membrane and form dimers with one another upon binding of growth factor from outside the cell (Fig. 13.4). Binding of growth factors evokes tyrosine-kinase activity from intracellular domains of the receptors and initiates proliferation signals through the cell cytoplasm.

EGFR is frequently overexpressed in solid tumors. Overexpression has been observed in a variety of tumor types. For this reason, the EGF receptor has been an attractive target for the design of therapeutic drugs. Monoclonal antibodies were developed to block ligand (growth factor) binding to the receptor. Agents were also designed to inhibit the kinase activity of the receptor. The efficacy of these **tyrosine-kinase inhibitors** (TKIs) has been confirmed in clinical trials.

IHC analysis of EGFR protein expression, similar to the testing for HER2/neu protein overexpression, was assessed for correlation with response to monoclonal antibody drugs[8] and methods approved by the U.S. Food and Drug Administration (FDA) for this application. Interpretation of the results of *EGFR* expression testing and the predictive value of the test are not always straightforward, however.[8,9] Quantitative polymerase chain reaction (PCR) has also been proposed to assess *EGFR* gene copy number through increased gene expression.[10,11]

Particular mutations in the kinase domain of the EGFR protein have proven better predictors of response to tyrosine-kinase inhibiting agents.[12–15] These mutations can be detected by a number of methods, including sequence-specific PCR, single-strand conformational polymorphism (SSCP), and direct sequencing.[13,16,17] The ever-expanding spectrum of clinically significant mutations has increased the use of next-generation sequencing (NGS) technology for this purpose.[18,19]

Testing for predictors of response or prognosis, even with comprehensive sequencing, is complex because multiple clinical and genetic factors contribute to the response to targeted therapies and the natural course of the tumor. These include intronic polymorphisms in the *EGFR* gene,[20] expression of other components of the signal transduction pathway,[21] or other tumor suppressors such as p53.[22] Also, multiple molecular and morphological characteristics can be found in single tumors as a result of tumor heterogeneity.[23,24]

Kirsten Rat Sarcoma Viral Oncogene Homolog, *K-ras* (12p12); Neuroblastoma ras, *N-ras* (1p13); and Harvey Rat Sarcoma Viral Oncogene Homolog, *H-ras* (11p15)

Signals from extracellular stimuli, such as growth factors or hormones, are transmitted through the cell cytoplasm to the nucleus, resulting in cell proliferation or differentiation (Fig. 13.5). The mitogen-activated protein kinase (MAPK) pathway is a cascade of phosphorylation events that transduces growth signals from the cell membrane to the nucleus. Critical components of this pathway are small proteins that bind to GTP in order to become active. These small GTP-binding proteins include the *ras* genes that receive signals from the cell surface proteins and activate the initial steps of the MAPK cascade (Fig. 13.6). Gain-of-function mutations in *ras* proto-oncogenes occur in many types of cancers.

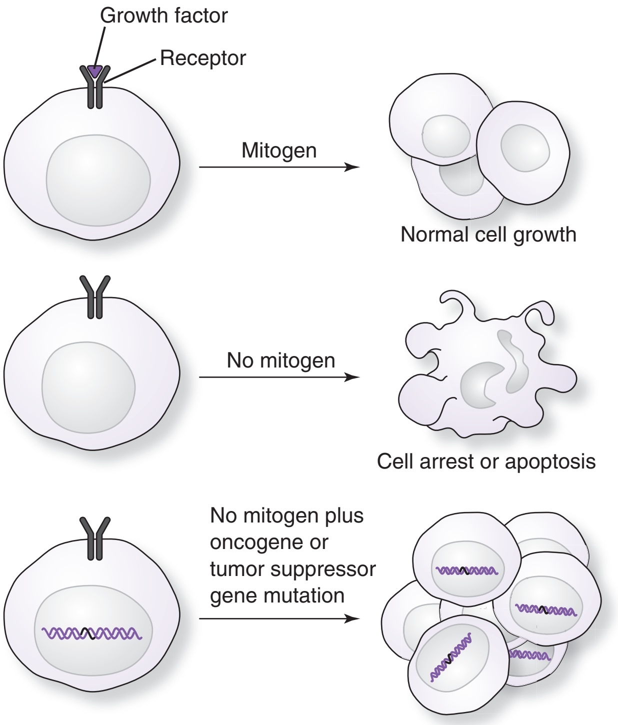


FIGURE 13.5 Normally, cells grow in the presence of nutrients and factors that stimulate cell division (mitogens). Lack of mitogen stimulation results in cell arrest, or apoptosis. If oncogene or tumor-suppressor gene mutations stimulate aberrant growth signals, cells grow in the absence of controlled stimulation.

Advanced Concepts

Regulation of ras GTPase activity is controlled by rasGAP. Several other GTPase-activating proteins (GAPs) besides rasGAP are important in **signal transduction.** Two clinically important proteins of the GAP family of proteins are the gene product of the neurofibromatosis type-1 *(NF1)* locus and the gene product of the breakpoint cluster region *(BCR)* gene. The *NF1* gene is a tumor-suppressor gene, and the protein encoded is called neurofibromin. The *BCR* locus is rearranged in the Philadelphia+ chromosome (Ph$^+$) observed in chronic myelogenous leukemias and acute lymphocytic leukemias.

Mammals have three different *ras* genes that produce four similar proteins, K-ras, N-ras, and H-ras. The Ras proteins differ only in their carboxy termini, the end of the proteins that anchor them to the inner surface of the cell membrane (Table 13.1 and Fig. 13.7). Because

TABLE 13.1 Four ras Proteins Synthesized From Three Genes, K-ras, N-ras, and H-ras*

Protein	Modification	Location
K-ras4A	Farnesylation + palmitoylation	?
K-ras4B	Farnesylation + polybasic amino acids	Plasma membrane
N-ras	Farnesylation + palmitoylation	Golgi
H-ras	Farnesylation + palmitoylation	Golgi

**K-ras*4A and *K-ras*4B arise from alternate splicing of transcripts of the same gene.

they bind and hydrolyze GTP for energy, the *ras* genes are members of a family of G-proteins. The GTP hydrolysis is regulated by GTPase-activating proteins (GAPs).

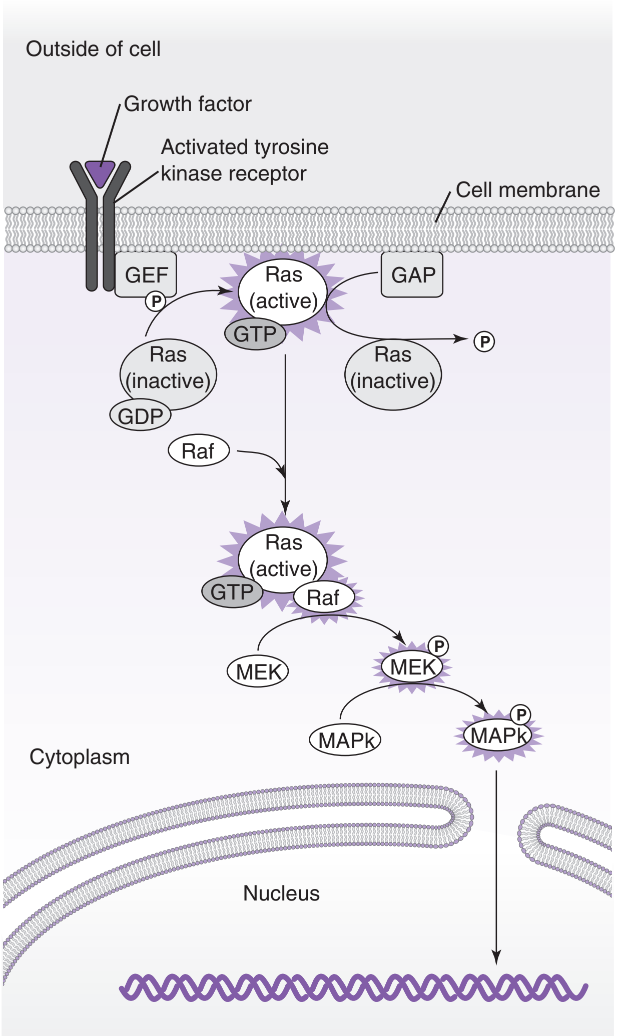


FIGURE 13.6 Activation of membrane-bound K-ras is initiated by activated receptors bound to mitogens (growth factors or hormones). Active K-ras bound to GTP then initiates a cascade of phosphorylation events that ends in the nucleus, where transcription factors modulate gene expression. GDP/GTP exchange on K-ras is modulated by GTPase-activating proteins (GAPs), guanosine nucleotide exchange factors (GEFs), and guanosine nucleotide dissociation inhibitors (GDIs).

Mutations in *K-ras* are the most common oncogene mutations in human cancers. The most frequently occurring mutations are located in codons 12, 13, 22, and 61 in exons 2 and 3 of the *KRAS* gene. Clinically significant mutations may also be found in exon 4. These mutations affect sequences coding for the GTP-binding domain of the protein and throw the KRAS protein into a permanently active state that does not require stimulation from GTP hydrolysis. As a result, the RAS proteins harboring these single-nucleotide substitutions remain constitutively active in the GTP-bound form.

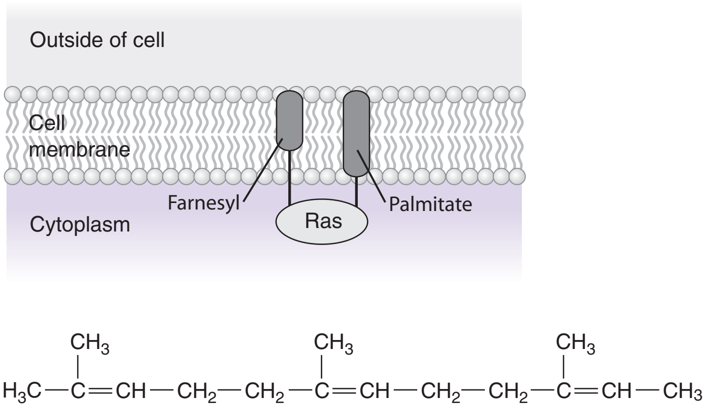
KRAS mutations are highly correlated with tumor histology and may predict the progress of tumorigenesis in early-stage tumors. Furthermore, the presence of *KRAS* mutations may affect treatment strategy, especially with targeted therapies such as kinase inhibitors and farnesyl-transferase inhibitors (KRAS protein is localized to the cell membrane through a farnesyl group). *KRAS* mutations are also a target of liquid biopsies, that is, analysis of tumor-specific mutations in blood.[25]

KRAS mutations are detected and identified by direct sequencing. Site-specific methods such as pyrosequencing have also been developed. Sequencing of selected exons of the *KRAS* and *NRAS* and downstream *BRAF* genes have become a common method for mutation analysis in lung, colon, thyroid, and skin cancer.[26,27] Certain characteristics of tumor tissue, such as poor cellularity, tumor heterogeneity, and previous chemotherapy treatments affecting tumor cells, confound the molecular diagnosis. These factors can be assessed and noted during initial tissue collection and processing.[28]

Ewing Sarcoma, *EWS* (22q12)

A group of tumors arising from primitive neuroectodermal tissue (PNET), Ewing sarcomas comprise a family of childhood neoplasms referred to as the Ewing family. Although immunohistochemical staining for the cell surface protein HBA71 (p30/p32MIC2), a neuron-specific enzyme, is helpful in the diagnosis of these tumors, no unique characteristics distinguish the different types of tumors that make up this group.

Detection of specific translocations by cytogenetic or molecular methods is useful for diagnostic and prognostic accuracy (Table 13.2). Translocations involving the *EWS* gene at 22q12 (also called *EWSR1* for *EWS* breakpoint region 1) with the *FLI-1* gene at 11q24, t(11;22)



$$\mathrm{H_3C-\overset{CH_3}{\overset{|}{C}}=CH-CH_2-CH_2-\overset{CH_3}{\overset{|}{C}}=CH-CH_2-CH_2-\overset{CH_3}{\overset{|}{C}}=CH-CH_3}$$

FIGURE 13.7 Ras proteins are anchored to the cell membrane through farnesyl groups *(chemical structure at bottom)* and palmitoyl groups on the ras proteins. K-ras has only farnesyl groups, whereas N-ras and H-ras have both farnesyl and palmitoyl groups.

TABLE 13.2 EWS Translocation Partners

Translocation	Tumor
EWS-FLI-1	Ewing sarcoma, peripheral PNET (72%)
EWS-ERG	Ewing sarcoma, peripheral PNET (11%)
EWS-WT1*	Desmoplastic small round cell tumor
EWS-ATF1	Clear cell sarcoma

*The *WT1* gene is also associated with Wilms' tumor (WT), one of the common solid tumors of childhood, accounting for 8% of childhood cancers. Several genes or chromosomal areas affect the development of WT: WT1 at 11p13, WT2 at 11p15.5, WT3 at16q, WT5 at 7p15-p11.2, and WT4 at 17q12-q21.

(q24;q12) are present in 85% of Ewing sarcomas. Another translocation between *EWS* and the *ERG* gene at 21q22 is present in 5% to 10% of Ewing sarcomas. Other partners for the *EWS* gene, such as *ETV1* at 7p22, *E1AF* at 17q12, and *FEV* at 2q33, are present in fewer than 1% of cases.[29] The occurrence of these rearrangements was first revealed by cytogenetics[30,31] and then by PCR methods.[32,33]

Laboratory testing at the molecular level includes detection of the tumor-specific translocations by reverse transcriptase polymerase chain reaction (RT-PCR; Fig. 13.8). Positive results are revealed by the presence of a PCR product. Negative specimens will not yield a product. As with all assays of this type, an amplification control, such as GAPDH or 18S RNA, must accompany all samples to avoid false-negative results. These tests can be performed on tissue or liquid biopsies.[34,35]

Synovial Sarcoma Translocation, Chromosome 18—Synovial Sarcoma Breakpoint 1 and 2, *SYT-SSX1, SYT-SSX2* t(X;18)(p11.2;q11.2)

A recurrent reciprocal translocation between chromosome 18 and the X chromosome is found in synovial sarcoma, a rare type of cancer of the muscle, fat, fibrous tissue, blood vessels, or other supporting tissue of the body. Synovial sarcoma accounts for 8% to 10% of all sarcomas and occurs mostly in young adults. About 80% of cases have the t(X;18) translocation.

The t(X;18) translocation fuses the synovial sarcoma translocation, chromosome 18 gene (*SS18* or *SYT*) with either of two related genes, synovial sarcoma translocated to X (*SSX1* and *SSX2*), on the X chromosome. There are five *SSX* variants, *SSX1, SSX2, SSX3, SSX4,* and *SSX5*. With rare exceptions, only *SSX1* and *SSX2* are fused to *SYT* in the t(X;18) translocation.[36] The fusion gene acts as an aberrant transcription factor, with both activation and repression functions from the *SYT* and *SSX* portions, respectively.

The t(X;18) translocation is detected by FISH or RT-PCR.[37] In the latter method, total RNA reverse-transcribed to cDNA is amplified with primers specific for *SSX* and *SYT* genes. In a semi-nested PCR version of this procedure, the *SSX* primer used in the first round is a consensus primer for both *SYT-SSX1* and *SYT-SSX2*. After the first amplification, *SSX1*- and *SSX2*-specific primers

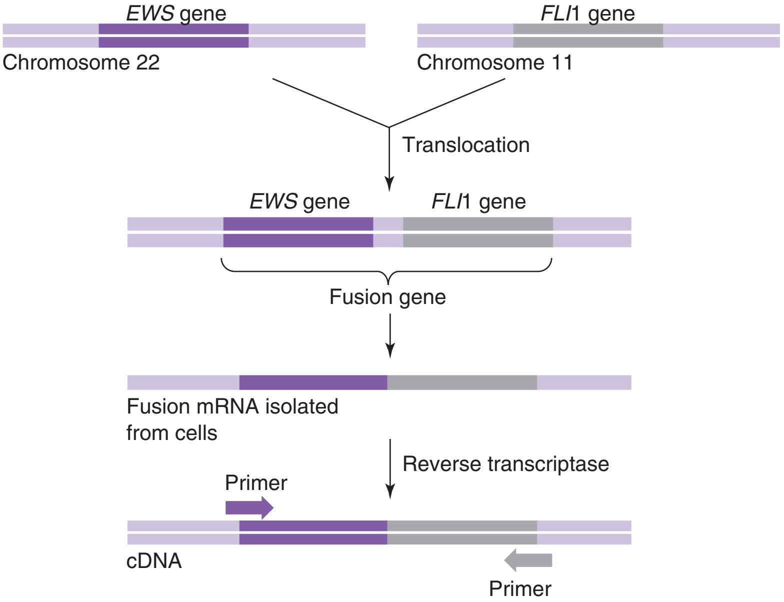


FIGURE 13.8 RT-PCR is used to detect the t(11;22) mutation in Ewing sarcoma. One primer is designed to hybridize to the *EWS* gene on chromosome 22 and one primer to the *FLI*1 gene on chromosome 11. If the translocation has occurred, the resulting fusion transcript isolated from the tumor cells will yield cDNA that is amplifiable with the *EWS* and *FLI*1 primers.

discriminate between the two translocation types. The PCR products are detected by agarose gel electrophoresis and ethidium bromide staining. This method can be performed on fresh, frozen, or fixed tissue, depending on the condition of the specimen RNA.

Paired Box-Forkhead in Rhabdomyosarcoma, *PAX3-FKHR*, *PAX7-FKHR*, t(1;13), t(2;13)

Rhabdomyosarcoma (RMS) is the most common soft tissue sarcoma of childhood, accounting for 10% of all solid tumors in children. In addition to alveolar rhabdomyosarcoma (ARMS), there are two additional histological forms of RMS: embryonal (RMS-E) and primitive (RMS-P). Although histological classification of RMS is sometimes difficult, accurate diagnosis is important for the management and treatment of this malignancy because ARMS has a worse prognosis than other subtypes.

Translocations involving the Forkhead in the Rhabdomyosarcoma gene (*FKHR,* also called *FOXO1A*) and the paired box genes (*PAX3* and *PAX7*) are frequently found in ARMS.[38] The chimeric genes resulting from the translocations encode transcriptional activators with DNA-binding motifs homologous to the forkhead transcription factor first discovered in the fruit fly, *Drosophila*. *PAX-FKHR* translocations have been observed in all subtypes of RMS but are more characteristic of ARMS. Furthermore, *PAX7-FKHR*, t(2;13), is associated with better outcomes than *PAX3-FKHR*, t(1;13). Mutations in the *PAX3* gene are also found in Waardenburg syndrome, a congenital auditory pigmentary syndrome.[39,40] The majority of ARMS displays the t(2;13) translocation, with the t(1;13) variant present with one-third the frequency as t(2;13). Both translocations are detected by FISH, RT-PCR, quantitative polymerase chain reaction (qPCR), and RNA sequencing.[41–43]

Tumor Protein 53, *TP53* (17p13)

Mutations in *TP53* are found in all types of cancer, and about 50% of all cancers have *TP53* mutations. The gene product of *TP53,* p53, is a 53,000-dalton DNA-binding protein that controls the expression of other genes. Normally, p53 participates in the arrest of cell division in the event of DNA damage. The arrest in the G1 phase of the cell cycle allows repair enzymes to correct the DNA damage before DNA synthesis begins. Once the damage is repaired, p53 protein is removed by binding to another protein, MDM2, and through degradation. When p53 is not functional, replication proceeds on damaged templates, resulting in the potential for further genetic abnormalities. Also, the mutant protein does not degrade properly and accumulates in the cell nucleus and cytoplasm. Due to the frequency and ubiquity of its mutations, mutated p53 protein is a potential therapeutic target.[44] Small molecules that can reactivate mutant p53 protein have been tested in clinical trials.

Several studies have shown that mutated *TP53* in tumor tissue is also an indicator of poor prognosis in breast cancer, lung cancer, colon cancer, leukemia, and other types of cancers. The significance of *TP53* status as a predictor of decreased survival time or tumor relapse, however, became controversial.[45] Part of the controversy arose from the different methods used to detect *TP53* mutations. A common method was the detection of the stabilized mutant protein by IHC. Several monoclonal antibodies directed at different epitopes in the p53 protein were used for this purpose. Because normal p53 protein is transient, significant staining (+2 or above on a scale of 0 to +4) of p53 is considered positive for the mutation. Use of microarrays or panel sequencing to screen expression of multiple genes along with *TP53* was proposed as a more accurate method for predicting survival than either IHC or mutation analysis of *TP53* alone. Methods that include sequencing of the entire *TP53* coding region on cDNA, in combination with IHC, is another accurate approach.

SSCP and direct sequencing of microdissected tumor tissue are other methods often used to detect *TP53* mutations. SSCP methods cover selected exons 5 to 8 or 4 to 9 of the *TP53* gene because these exons encode the regions involved in DNA binding and protein–protein interactive functions of the p53 protein. Sequencing methods can include the entire gene. In early studies, mutations detected by IHC were not always consistent with mutations found by direct DNA analysis.[46,47] There are several explanations for these discrepancies (Table 13.3).

TABLE 13.3 Sources of Potential Inaccuracies of p53 Mutation Analysis

Method	False Positive	False Negative
IHC	Staining of normal p53 protein	Deletions or mutations in p53 that remove Ab binding epitopes; promoter mutations
SSCP	Alternate conformers; silent DNA polymorphisms	Less than 5% mutant cells in specimen; mutations outside of the exons screened
Sequencing	PCR mutagenesis; high background	Less than 10% mutant cells in specimen; mutations outside of sequenced area

In addition to screening for somatic alterations, mutation analysis of *TP53* is also performed to aid in the diagnosis of Li–Fraumeni syndrome, a cancer-prone condition caused by inherited mutations in the p53 gene. In this case, normal tissue will be heterozygous for the mutation, removing the challenge of isolating pure samples of tumor tissue. Once an inherited mutation is detected, further analysis of relatives requires targeting only that mutation.

Ataxia Telangiectasia Mutated Gene, *ATM* (11q22)

Predisposition to cancer is one symptom of the neurological disease ataxia telangiectasia (AT). AT occurs in at least 1/40,000 live births. This disease is caused by mutations in the *ATM* (A-T mutated) gene on chromosome 11. *ATM* mutations are also present in some types of leukemias and lymphomas. Carriers of the autosomal-recessive mutations in *ATM* are at increased risk for developing leukemia, lymphoma, or other types of cancers.

The *ATM* gene product is a member of the phosphatidylinositol-3 kinase family of proteins that respond to DNA damage by phosphorylating other proteins involved in DNA repair and/or control of the cell cycle. The ATM protein participates in pausing the cell cycle at the G1 or G2 phase to allow completion of DNA repair.

Direct DNA sequencing is the method of choice for detection of *ATM* mutations, especially in family members of carriers of previously identified mutations. Other methods, such as SSCP, have been used. A functional test for the repair of double-strand breaks induced by irradiation was developed for *ATM*. For this assay, exponentially growing cells were irradiated, followed by the addition of Colcemid to inhibit spindle formation. The cells were harvested for Giemsa staining, and the karyotypes were examined. The ratio of aberrations/cell was calculated from the number of chromatid and chromosome breaks (counted as one breakage event) in addition to dicentric chromosomes, translocations, ring chromosomes, and chromatid exchange figures (counted as two breakage events).[48] More recent measures of DNA damage utilize chromogenic reporters.[49,50]

If a mutation is identified in a patient with ataxia telangiectasia, other family members may be tested for the presence of the same mutation. Presence of a mutation in family members identifies those with an increased risk of AT. Heterozygous carriers of an *ATM* mutation may also be at increased risk for mantle cell lymphoma, B-cell lymphocytic leukemia, or T-cell prolymphocytic leukemia.

Breast Cancer 1 Gene, *BRCA1* (17q21), and Breast Cancer 2 Gene, *BRCA2* (13q12)

Approximately 5% of breast cancers result from inherited gene mutations, mostly in the breast cancer genes *BRCA1*[51,52] and *BRCA2*.[53,54] Women who carry a mutation in *BRCA1* have a 60% to 80% lifetime risk of breast or ovarian cancer. Men carrying a mutation, especially in *BRCA2,* have a 100-fold increased risk of breast cancer compared with men without a mutation, as well as increased risk of colon and prostate cancer. Both men and women can transmit the mutation to subsequent generations.

The *BRCA1* gene product has a role in embryonic development, and both *BRCA1* and *BRCA2* gene products are involved in DNA double-strand break repair by homologous recombination. Defects in homologous recombination repair (HRR) are associated with mutations in repair genes, including *BRCA1/BRCA2, ATM, ATR, PALB2, RAD51, CHEK1,* and *CHEK2,* as well as loss of *BRCA1* expression through promoter methylation or overexpression of the *BRCA2* transcriptional repressor EMSY.[55] Poly(ADP-ribosyl)transferase or PARP is also part of the repair process, and therapeutic agents inhibiting its activity in combination with other repair gene inhibitors have shown efficacy against *BRCA*-associated cancers.[56]

Advanced Concepts

PolyADP-ribosyl transferase modifies nuclear proteins by polyADP-ribosylation. The modification is involved in the regulation of important cellular processes, including the molecular events involved in the recovery of cells from DNA damage. In addition, this enzyme may be the site of mutation in Fanconi anemia and may participate in the pathophysiology of type 1 diabetes. Fanconi anemia is a genetically heterogeneous recessive disorder characterized by defective DNA repair. Mutations in at least 15 genes can cause Fanconi anemia. The Fanconi anemia complementation group (FANC) includes FANCA, FANCB, FANCC, FANCD1 (=BRCA2), FANCD2, FANCE, FANCF, FANCG, FANCI, FANCJ, FANCL, FANCM, and FANCN (=PALB2). These proteins are related as components of the nuclear protein complex.

SSCP, **protein truncation tests,** chromosome breakage tests, and other procedures have been used to screen for mutations in the *BRCA* genes. The method used for clinical applications, however, is direct sequencing. Three mutations, 187delAG and 5382insC in *BRCA1* and 6174delT in *BRCA2,* occur frequently in particular ethnic populations. These known mutations can be detected by several targeted assays, including sequence-specific PCR and allele-specific oligomer hybridization.

Just as with any genetic analysis, testing for *BRCA1* and *BRCA2* mutations requires patient counseling and education. The significance of a *BRCA* mutation test will depend on several factors, including penetrance of the gene mutations. Furthermore, if a mutation is not detected in the coding sequences of the genes, the possibility of mutations in the noncoding regions cannot be ruled out.

Von Hippel–Lindau Gene, *VHL* (3p26)

Benign blood vessel tumors in the retina were first reported by Eugen von Hippel, a German ophthalmologist, in 1895. In 1926, Arvid Lindau, a Swedish pathologist, further noted that these retinal tumors were linked to tumors in the blood vessels in other parts of the central nervous system, sometimes accompanied by cysts in the kidneys and other internal organs, and that the condition was heritable. The Von Hippel–Lindau syndrome (VHL) is now recognized as a genetic condition involving the abnormal growth of blood vessels in organs, especially those that are particularly rich in blood vessels. It is caused by mutations in the *VHL* gene, which is located on the short arm of chromosome 3. Normally, *VHL* functions as a tumor-suppressor gene, promoting cell differentiation. VHL syndrome is a predisposition for renal cell carcinoma and other cancers.[57]

Deletions, point mutations, and splice-site mutations have been described in patients with VHL. In addition, cases of renal cell carcinoma and tumors of the adrenal gland are accompanied by somatic mutations in the *VHL* gene. Direct sequencing is the preferred method of testing for *VHL* gene mutations.

V-myc Avian Myelocytomatosis Viral-Related Oncogene, Neuroblastoma-Derived, *MYCN* or *n-myc* (2p24)

Myc family proteins, activated by mitogen signals, increase the expression of several genes through interactions with conserved cis elements (Myc-boxes) and transcriptional coactivators. The Myc family includes the genes *c-myc, n-myc,* and *l-myc.* They are basic helix-loop-helix transcription factors.[58]

The *c-myc* oncogene (8q24.21) is the most frequently expressed myc. The *c-myc* gene is amplified in breast and ovarian cancer, lymphomas, and leukemias. *Myc* expression is regulated by transcription factors specific to each cancer type. Increased *Myc* expression is a negative prognostic marker. Expression of *l-myc* (1p34.2) can induce differentiation in cultured cells.[59] Amplification of *l-myc* is associated with oral cancer.

The *n-myc* gene on the short arm of chromosome 2 (2p24) is amplified in cases of neuroblastoma and retinoblastoma. *n-myc* is an oncogene that is counteracted by the tumor-suppressor gene neurofibromatosis type 1 *(NF1)*. *Myc* gene amplification (or protein expression) is detectable by IHC, FISH, sequencing, or array analysis.[60] Transcription of *n-myc* may also be measured using qPCR.[61]

V-Ros Avian UR2 Sarcoma Virus Oncogene Homolog 1 *(ROS1)* Proto-Oncogene (6q22.1) and Rearranged During Transfection *(RET)* Proto-Oncogene (10q11)

The *ROS1* oncogene, coding for a membrane receptor tyrosine kinase, is rearranged in a variety of human cancers. The resulting fusion protein contains a constitutively active ROS1 kinase domain and drives cellular proliferation. *ROS1* is rearranged in 1% to 3% lung adenocarcinomas.

The *RET* proto-oncogene is located on the long arm of chromosome 10 (10q11). Its gene product is a membrane tyrosine kinase that participates in sending cell growth and proliferation signals to the nucleus. The first intron in the gene covers about 24 kb, with exons 2 to 20 contained in the remaining 31 kb. This general structure of a large first intron with small exons is characteristic of tyrosine-kinase receptors, such as the receptors for the *KIT, EGFR,* and platelet-derived growth factor *(PDGF)* genes.

The *RET* gene is an example of how different mutations in the same gene result in different diseases. Translocations that result in overexpression of *RET* are found in thyroid papillary carcinomas. Point mutations that activate *RET* (also called *MEN2A*) are found in inherited multiple endocrine neoplasia (MEN) syndromes, a group of diseases resulting in abnormal growth and function of the pituitary, thyroid, parathyroid, and adrenal glands. In contrast, loss-of-function mutations in the *RET* gene are found in Hirschsprung disease, a rare congenital lack of development of nerve cells in the colon that results in colonic obstruction. Mutations have been reported in about 50% of congenital cases and 20% of sporadic cases of this disorder. Because about 16% of children with congenital central hypoventilation syndrome (CCHS) have Hirschsprung disease, *RET* mutations were also sought in CCHS, but most of the mutations detected were determined as polymorphic variants.[62]

Detection of *RET* gene mutations can aid in the diagnosis of MEN diseases. Clinical testing targets, mainly exons 10, 11, and 16, are where most reported mutations have been found.

RET and *ROS1* rearrangements are detected by IHC, FISH, or sequencing.[63] *RET* alterations occur in about 1% of lung cancers, and both RET and ROS1 are therapeutic targets.[64]

Anaplastic Lymphoma Receptor Tyrosine Kinase *(ALK)* Proto-Oncogene, 2p23.1

ALK is a receptor tyrosine kinase, classified in the insulin receptor superfamily. The *ALK* gene has been found to be rearranged, mutated, or amplified in tumors of various types, including lymphomas, neuroblastoma, and non–small-cell lung cancer. Chromosomal rearrangements are the most common genetic alterations in this gene, with multiple fusion gene partners, including the Echinoderm Microtubule-Associated Protein-Like 4 gene (*EML4*; 2p21) and others. *ALK* gene rearrangements are detected by FISH analysis and sequencing.

V-Kit Hardy-Zuckerman 4 Feline Sarcoma Viral Oncogene Homolog, *KIT, c-KIT* (4q12)

KIT protein is a transmembrane receptor with tyrosine kinase activity. Mutations in this gene are associated with gastrointestinal stromal tumors (GISTs), mast cell disease, and AML. KIT activation has oncogenic activity. Targeted therapeutics have been used to treat patients with melanoma and GIST. Increased KIT activity can result from amplification, overexpression or missense mutations. IHC is used to detect increased KIT protein, and sequencing is used for detection of the missense mutations.

Other Molecular Abnormalities

Increasing numbers of molecular abnormalities are being used to aid in the diagnosis and monitoring of solid tumors. Some examples of potential diagnostic targets are shown in Table 13.4. As molecular aberrations in oncogenes and tumor-suppressor genes are identified, molecular analysis becomes more important in their rapid and accurate detection.

Microsatellite Instability

Lynch syndrome, or hereditary nonpolyposis colorectal cancer (HNPCC),[65] is an inherited cancer predisposition syndrome, accounting for 3% to 5% of colon cancers and 3% of endometrial cancers. There is also an increased risk of developing other types of cancers, including cancer of the stomach, breast, ovary, pancreas, prostate, urinary tract, liver, kidney, or bile duct. Predisposition to cancer in this syndrome is caused by mutations in the mismatch repair (MMR) protein complexes, including mutS homologs, *MSH2, MSH6,* mutL homolog *MLH1,* and the human postmeiotic segregation *hPMS2*.[66] These complexes are responsible for correcting replicative errors and mismatched bases in DNA. The MMR system was originally discovered in bacteria *(Escherichia coli)* and further studied in yeast *(Saccharomyces cerevisiae)*. Similar (homologous) genes were subsequently identified in humans and named after the bacterial and yeast genes (Table 13.5). The protein complex binds to mismatched bases in the DNA double helix or loops formed by replicative errors. The repair is methyl-directed. At the end of S phase (DNA replication), the system recognizes errors in the newly synthesized (unmethylated) daughter strand and uses the template strand, which is methylated, as a guide for repair (Fig. 13.9).

Included in the types of DNA lesions that are repaired by this system are replication errors (RERs) caused by slippage between the replication apparatus and the DNA template (Fig. 13.10). RERs occur especially in microsatellites where one to three nucleotides are repeated in the DNA sequence. If the errors remain in the DNA until the next round of replication, new alleles will arise, generating increasing numbers of alleles for the locus, or **microsatellite instability (MSI).** In contrast, a stable locus will retain the same alleles through many rounds of replication. Microsatellite slippage occurs about every 1,000 to 10,000 normal cell divisions, most of which are repaired in normal cells. Dysfunction of one or more components of the MMR system will result in MSI, an increase in the number of alleles due to loss of repair. The majority of MMR mutations in Lynch syndrome are found in the *MSH2* and *MLH1* genes. Mutations in *hPMS2* account for 9% of MMR variants, compared with 39% *MLH1* and 33% *MSH2* and 19% *MSH6* (see http://www.insight-database.org/genes). Mutations in *hPMS1* and *MSH3* are rare. Although direct sequencing of the affected genes is definitive and identifies the specific mutation in a family, the test may miss mutations outside of the structural gene sequences or in other genes. Screening for MSI normal MMR proteins by IHC is often the first step in MSI analysis.

Lack of staining of the normal protein is evidence of loss of MMR function. Because this loss of MMR gene function causes MSI, MSI can be used to screen indirectly for mutations in the MMR genes. With an inherited

TABLE 13.4 Molecular Abnormalities in Some Solid Tumors

Gene	Location	Mutation	Detection Method	Associated Disease
Adenomatous polyposis of the colon	5q21	5q deletion, t(5;10)	Southern blot, FISH, sequencing, SSCP	Familial adenomatous polyposis of the colon
Retinoblastoma *(Rb, RB1)*	13q14.1	13q deletion, t(X;13)	Southern blot, FISH, sequencing	Retinal neoplasm, osteosarcoma
MET proto-oncogene, hepatocyte growth factor receptor	7q31	Missense mutations	Sequencing	Renal carcinoma
KIT proto-oncogene, stem cell factor receptor *(SCFR)*	4q12	Missense mutations	Sequencing	Gastrointestinal stromal tumors
Folliculin *(FLCL, BHD)*	17p11.2	Insertions, deletions in a C8 tract in exon 11	Sequencing	Birt–Hogg–Dube syndrome (hair follicle hamartomas, kidney tumors)
Fumarate hydratase	1q42.1	Frameshift mutations	Sequencing	Hereditary leiomyomatosis, renal cell cancer
Activin A receptor *(ALK4)*	12q13	Gene fusion		Pancreatic carcinoma, lung cancer
Epidermal growth factor receptor *(EGFR)*	7p12	Missense mutations, deletions	SSP-PCR, SSCP, Sequencing	Lung cancer, glioblastoma
Avian erythroblastic leukemia viral oncogene homolog *(ERBB-2, HER2)*	17q21	Amplification	Immunohistochemistry, FISH	Breast cancer
Platelet-derived growth factor alpha *(PDGFRa)*	4q12	Deletion	Immunohistochemistry, FISH, CISH	Idiopathic hypereosinophilic syndrome

TABLE 13.5 Genes of the MMR System

Human Gene	Bacterial Gene	Function
MSH2	*MutS*	Single mismatch, loop repair
MSH3	*MutS*	Loop repair
MSH4	*MutS*	Meiosis
MSH5	*MutS*	Meiosis
MSH6/GTBP	*MutS*	Single mismatch repair
MLH1	*MutL*	Mismatch repair
hPMS2	*MutL*	Mismatch repair (postmeiotic segregation in yeast)
hPMS1	*MutL*	Mismatch repair (postmeiotic segregation in yeast)

mutation in one copy of an MMR gene, somatic mutation of the remaining copy will result in the MSI phenotype in tumor cells. MSI, therefore, will be apparent in the tumor where both functional copies of the gene have been lost but not in normal tissue that retains one normal copy of the gene. To perform MSI analysis, therefore, normal and tumor tissue from the patient must be compared. MSI is apparent from the increased number of alleles in the tumor tissue compared with that in the normal tissue after PCR amplification of microsatellite loci and gel electrophoresis (Fig. 13.11) or **capillary gel electrophoresis** (Fig. 13.12). The detection of instability (more bands or peaks in the tumor tissue compared with the normal tissue) is strong evidence for Lynch syndrome. The National Cancer Institute has recommended that screening two mononucleotide-repeat loci, BAT25 and BAT26, and three dinucleotide-repeat loci, D5S346, D2S123, and D17S250, is sufficient for determination of MSI.[67] Alternate markers have been proposed, and some laboratories test additional loci to ensure amplification of at least five loci and address discordant or borderline samples.[68] Further, mononucleotide-repeat structures may be more sensitive markers for MSI than dinucleotide repeats, so some laboratories prefer to use only mononucleotide-repeat loci.

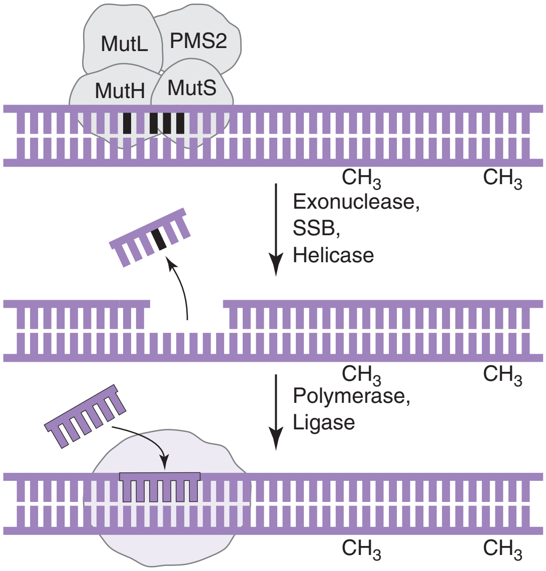


FIGURE 13.9 The mismatch repair system recognizes a mismatch *(top)* in the newly synthesized unmethylated strand of DNA. The complex of proteins recruits exonucleases, single-strand DNA-binding proteins (SSB), and helicases to remove the erroneous base *(center)*. Polymerases and DNA ligase then replace the missing bases.

If at least two of the five, or if more then 30% of loci show instability, the specimen is classified as having high instability (MSI-H). Tumors showing MSI in one or a minority of loci tested are classified as low instability (MSI-L). If no MSI is detected in the loci tested, the tumor is stable (MSS). Initially, MSI-L and MSS tumors are interpreted as microsatellite stable, and MSI-H tumors are considered microsatellite unstable. More recent clinical correlations suggest that MSI-L may have specific implications.[69] MSI-H is reported as MSI unstable, with an increased likelihood that the patient has Lynch syndrome.[69] If MSI is discovered, the inherited mutation can be confirmed by direct sequencing of the MMR genes.

Sequencing will not detect epigenetic silencing of the *MLH1* gene, which can also result in MSI. A separate test for MLH1 promoter methylation by bisulfite sequencing

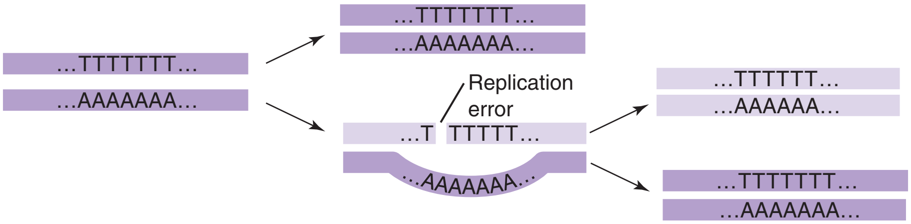


FIGURE 13.10 Replication errors result from slippage during DNA replication. If the error is not repaired, the next round of replication will create a new allele *(top right)* of the original locus. Additional uncorrected errors will produce more alleles.

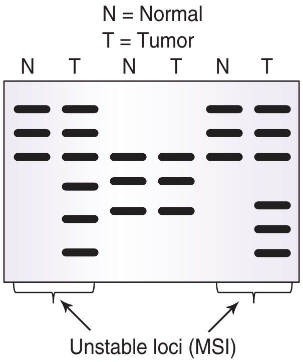


FIGURE 13.11 Microsatellite instability (MSI) is detected by increased alleles compared with stability at the same locus.

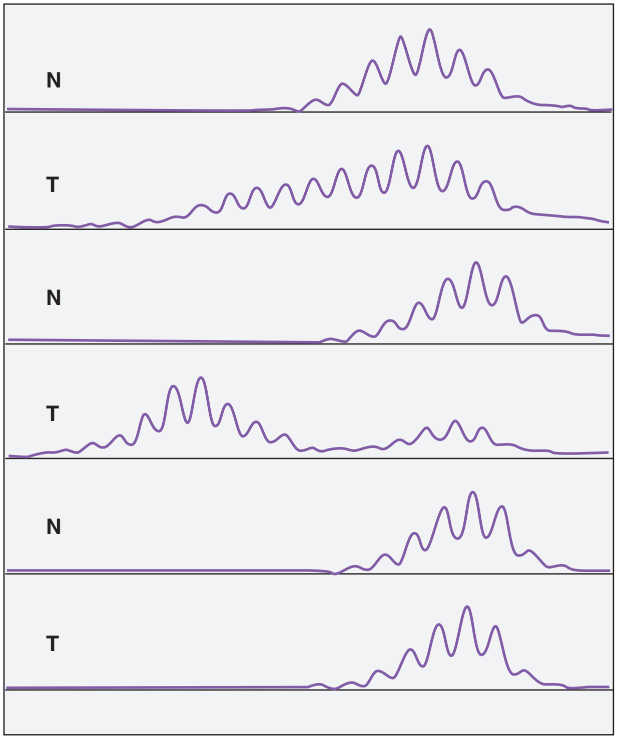


FIGURE 13.12 MSI detected by capillary gel electrophoresis. DNA from the tumor (T) is compared with DNA from normal cells from the same patient (N). Increased alleles in the tumor scans reveal instability at those loci *(top four scans)*. Stable loci look the same in normal and tumor tissue *(bottom two scans)*.

or methylation-specific PCR may also be performed.[70] Testing for deletion of the *EPCAM* gene by sequencing or multiplex ligation-dependent probe amplification (MLPA) is another approach to assess linked deletion of MMR loci.[71] The choice of methodology used will vary among cases. The cost of testing increases with IHC, PCR, and sequencing, respectively.[72]

Loss of Heterozygosity

Increased risk of cancer is a concern for members of families carrying cancer-predisposition mutations. These mutations are recessive regarding the cancer or malignant cell phenotype but dominant regarding the cancer risk. Therefore, relatives of patients with suspected inherited conditions are tested for purposes of cancer surveillance and family planning.

Once the familial gene mutation is identified in the patient or **proband,** targeted tests for the mutation in other family members are performed. In an inherited condition such as Lynch syndrome or a history of breast and ovarian cancer (HBOC), blood samples are sufficient for mutation analysis. In tumor cells, further testing for **loss of heterozygosity (LOH)** may be performed. LOH reveals the loss of the "good allele" at a locus, uncovering the homologous locus with a recessive mutation. LOH can be detected by PCR amplification of heterozygous STR or variable-number tandem repeat (VNTR) loci closely linked to the disease gene.[73] Amplification of loci in tumor cells with LOH will reveal a loss of the allele linked to the normal allele of the gene when compared with the mutant allele (Fig. 13.13). Comparing peak heights in normal (N) and tumor (T) tissues, the formula for LOH is as follows:

$$\frac{(\text{peak height of normal allele in N}/\ \text{peak height of normal allele in T})}{(\text{peak height of mutant allele in T}/\ \text{peak height of mutant allele in N})}$$

A ratio of less than 0.5 or more than 2 indicates LOH. Alternately, LOH is assumed if the peak height of the normal allele in the tumor is less than 40% of the height of the normal allele in the normal DNA. Technical artifacts such as allelic dropout or PCR bias complicate the interpretation. When analyzing tumor tissue, it may be necessary to test more than one area of a tumor to confirm LOH.

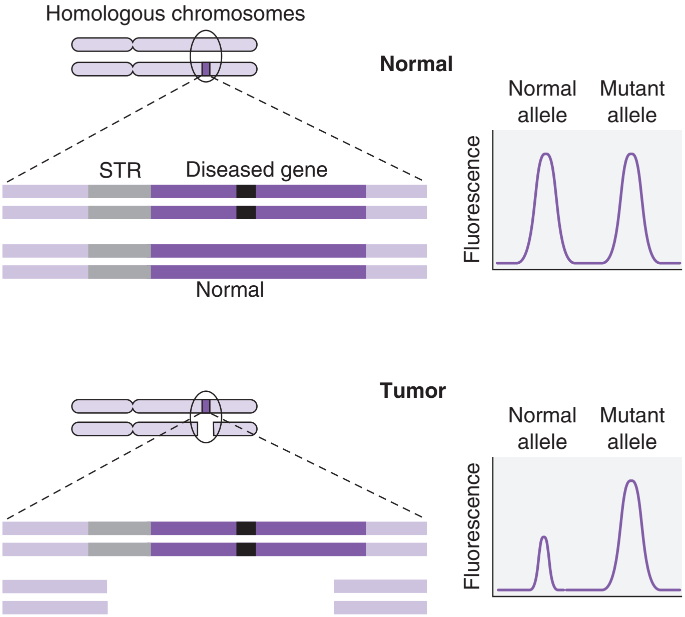


FIGURE 13.13 Loss of heterozygosity (LOH) is detected by PCR and capillary electrophoresis of heterozygous STR loci linked to disease genes. A deletion or loss of the normal allele uncovering a recessive mutant allele is identified by the loss of the STR linked to the normal gene *(right)*.

Liquid Biopsy

Solid tumors release cells and nucleic acids into circulation. This may have a role in metastatic cancers, which grow in multiple body sites. As the application of molecular testing advances, analysis of tumor cells has become increasingly informative. Surgical biopsies are commonly used for molecular analyses and are preferable to surgical resections in some cases. With the development of efficient nucleic acid isolation methods and highly sensitive analyses, tumor cells and DNA in body fluids have become a valuable source of information without the cost and comorbidity of surgical biopsies. The liquid biopsy is usually from blood plasma or serum or urine but may also be done on other body fluids. There are two approaches to liquid biopsies: cell free or circulating free nucleic acids (cfDNA or cfRNA), exosomes (cell-derived vesicles carrying nucleic acid or other molecules) and circulating tumor cells (CTC).

Circulating nucleic acid analysis can be classified as targeted or untargeted.[74] Targeted procedures are directed at specific mutations using sequencing or PCR methods specifically designed for those mutations or rearrangements. These methods can reach high sensitivity (1% to 0.001%) and provide an estimation of mutant allele frequency. Untargeted methods assess whole exomes or, more frequently, gene panels by sequencing and comparative genomic hybridization (CGH) array technology (e.g., digital karyotyping), with sensitivities of 0.2% to 2%. They can detect copy-number variations and structural abnormalities. Plasma DNA is the most frequent source material. Plasma RNA has seen less clinical application but has the potential to provide additional information on tumor gene expression and tissue-specific gene expression. The latter might be used to determine the tissue of origin of tumors as well.[75]

Urine has also been a source of genetic information from cfDNA. Urine analysis has been applied to the detection of acquired mutations in lung cancer after treatment with targeted therapies.[76,77]

Circulating tumor cells (CTCs), first reported in 1869, were considered to be a mechanism of tumor metastases from one tissue site to another. Although low in number, they can travel in clusters of 2 to more than 50 cells.[78] There are several technologies for the isolation of CTCs. Antibody capture, using immobilized antibodies to epithelial cell–specific membrane proteins, such as the epithelial cell adhesion molecule (EpCAM), is most frequently used. Alternatively, negative depletion of leukocytes can also be performed. Cell morphology and physical characteristics such as size (epithelial cancer cells are larger than leukocytes) or density may be used for selection. Once the tumor cells are enriched, they may be stained and other characteristics measured.

Extracted DNA is tested using allele-specific PCR or other sensitive methods. Gene-expression profiling and translocations may be found in extracted cDNA made from extracted RNA.

cfDNA and CTC are also sources of epigenetic information. Methylation patterns specific to tissues or tumor state can be assessed using sequencing, methylation-specific PCR, and methylation arrays. miRNA can be isolated from plasma as well. Although not yet part of routine clinical testing, the presence, absence, or expression levels of particular miRNA species may provide information on prognosis or tumor state. Just as with DNA variants, annotated databases will provide a basis for interpretation of the significance of epigenetic analyses.

MOLECULAR ANALYSIS OF LEUKEMIA AND LYMPHOMA

Gene Rearrangements

Gene rearrangements analyzed for hematological malignancies include **V(D)J recombination,** the normal intrachromosomal rearrangements in B and T lymphocytes, and the abnormal interchromosomal translocations that can occur in any cell type.

V(D)J Recombination

To enhance antibody diversity, lymphocytes undergo normal genetic rearrangement of immunoglobulin (Ig) heavy- and light-chain genes and T-cell receptor genes (Fig. 13.14). The **gene rearrangement** process is a series of intrachromosomal recombination events mediated by recombinase enzymes that recognize specific sequences flanking the gene segments. This process occurs independently in each lymphocyte, so that a diverse repertoire of antibodies is available to match any random invading antigen.

Immunoglobulin Heavy-Chain Gene Rearrangement in B Cells

Each antibody consists of two heavy chains and two light chains. The gene encoding the immunoglobulin heavy chain is located on chromosome 14. The unrearranged or **germline** configuration of the immunoglobulin

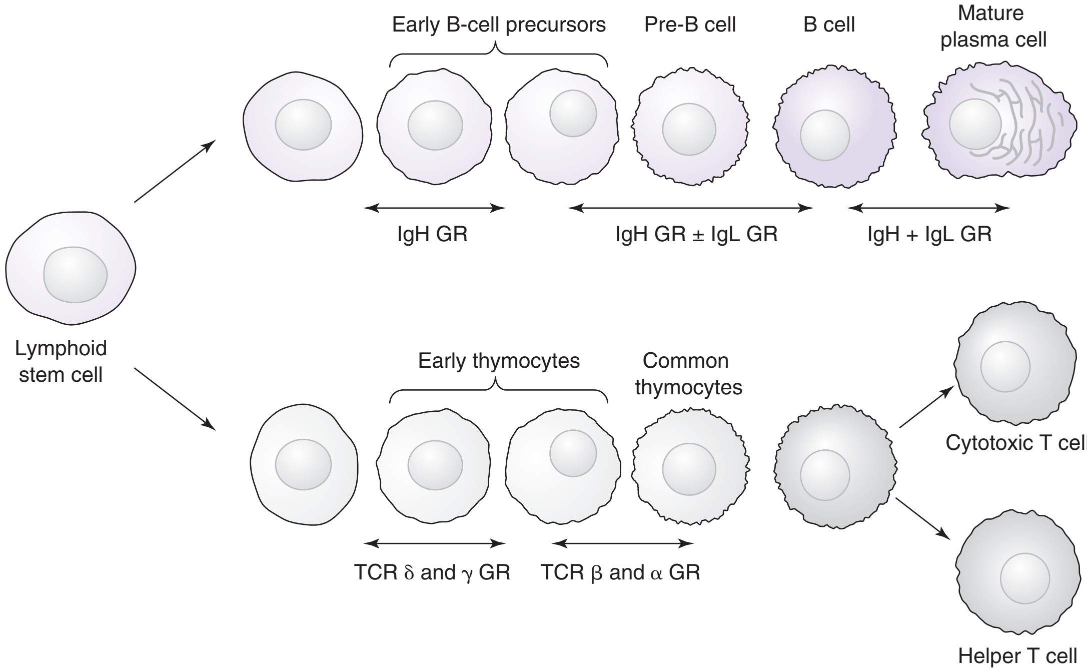


FIGURE 13.14 Gene rearrangements (GRs) are normal processes that occur in B and T lymphocytes as they mature from lymphoid stem cells. The genes coding for immunoglobulin heavy and light chains (IgH and IgL, respectively) begin the rearrangement process in early B cells and pre–B cells. The T-cell receptor (TCR) genes rearrange in the order δ, γ, β, and α chains.

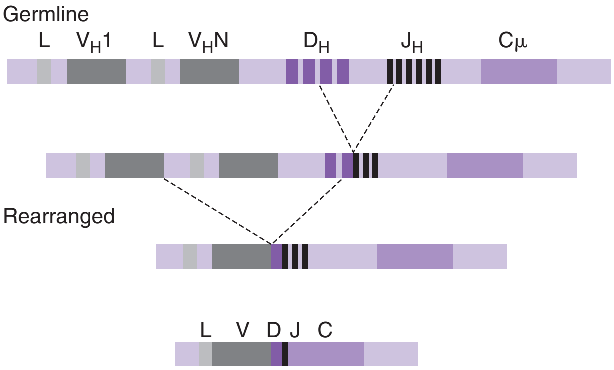


FIGURE 13.15 The immunoglobulin heavy-chain gene on chromosome 14 consists of a series of variable (V), diversity (D), and joining (J) gene segments (germline configuration). The V segments are accompanied by a short leader region (L). One of each type of segment, V, D, and J, is selected and combined by an intrachromosomal recombination event, first D and J, and then V and D. The C (constant) segments are joined through splicing or a secondary recombination event, class switching.

heavy-chain locus consists of a series of gene segments or repeated exons coding for the functional parts of the antibody protein (Fig. 13.15). These include 123 to 129 variable (V_H) regions (38 to 46 functional gene segments) and 9 joining (J_H) regions (6 functional), one of which will connect one variable region with a constant (C_H) region of the antibody or receptor. There are 11 constant regions (9 functional). The immunoglobulin heavy-chain gene also contains 27 diversity (D_H) regions (23 functional), one of which will connect the variable and joining regions. The V segments are each preceded by a leader region (L). The leader region codes for a short sequence of amino acids on the amino terminus of the protein that marks the antibody for secretion or membrane insertion.

As B lymphocytes mature, selected gene segments are joined together so that the rearranged gene contains only one of each V_H, D_H, and J_H segment (see Fig. 13.15). Initially, one D_H and one J_H segment are joined together. The DNA between the two segments is looped out and lost. The D_H–J_H rearrangement occurs in both alleles of the heavy-chain gene locus on both chromosomes. Then in only one allele, a V_H segment is chosen and joined to the D_H segment. The completion of the gene-rearrangement process on only one of the two immunoglobulin heavy-chain gene alleles is referred to as **allelic exclusion.** The rearrangement on the other chromosome will proceed if the first rearrangement fails or is not productive. The recombination events of the gene rearrangement are mediated by recombination activating genes, *RAG1* and *RAG2*. These genes code for enzymes that recognize short recombination signal sequences in the DNA, where they form a complex that initiates the cutting and re-ligation of the DNA.

When the DNA is cut in the process of gene rearrangement, terminal deoxynucleotidyl transferase may add nucleotides at the V–D–J junctions, further diversifying the coding sequences of individual antibody genes. After the V(D)J rearrangement occurs, another enzyme, activation-induced deaminase (AID), further changes the DNA sequence of the variable region. These variable regions are said to have undergone **somatic hypermutation.**

Advanced Concepts

Chronic lymphocytic leukemia (CLL) arising from cells harboring the "mutated" variable region (somatic hypermutation) have better prognoses than CLL arising from cells before the mutation process. Variable regions are sequenced as a clinical test in CLL. More than 2% sequence divergence from an "unmutated" reference sequence is considered "mutated," with a better prognosis than "unmutated" variable regions.

After the gene is transcribed, one of the constant regions is joined to the final messenger RNA by splicing or, alternatively, by a secondary recombination event **(class switching).** The maintenance of the constant regions in the DNA allows for antibody-type switching during the immune response.

Advanced Concepts

The constant region determines the **isotype** of the antibody IgM, IgD, IgG, IgE, or IgA. Each cell can make only one heavy-chain protein, although the isotype of the heavy chain may change. A mature

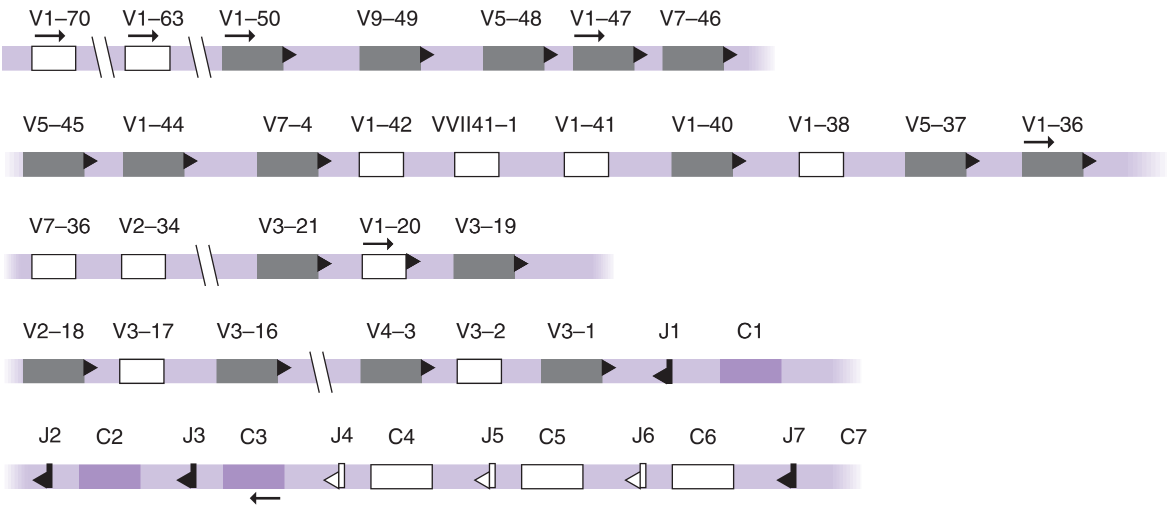


FIGURE 13.16 Immunoglobulin kappa and lambda light-chain loci consist of gene segments for variable (V), joining (J), and constant (C) regions. The variable regions are classified into sequence-related families (V1 to Vn). Each member of the family is given a number; for example, V7-4 is the fourth member of the V7 sequence family. Some of the gene segments are nonfunctional (open boxes). Recombination sites (triangles) are juxtaposed to each gene segment. Arrows denote primer-binding sites for PCR clonality testing.

B cell will initially produce IgD and some membrane IgM that will migrate to the cell surface to act as the antigen receptor. Upon antigen stimulation, the B cell will differentiate into a plasma cell expressing large amounts of secreted IgM. Some cells will undergo a **class-switch recombination,** placing the *VDJ* gene next to the genes encoding the IgG, IgE, or IgA constant regions. The B cells will express a different isotype during the secondary response. Most commonly, IgM (primary response) gives way to IgG (secondary response). Production of IgE or IgA instead of IgG can also occur. Class switching is mediated by different recombinase enzymes than those responsible for VDJ recombination.

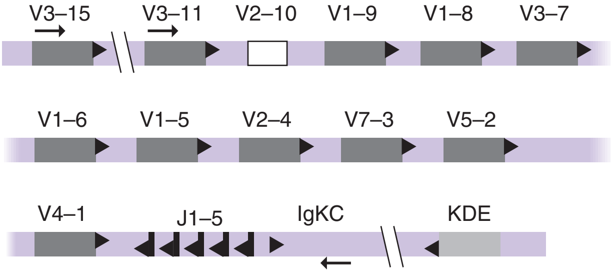


FIGURE 13.17 Light-chain genes are rearranged in a kappa-before-lambda order. The rearrangement of the kappa-deleting element (KDE) eliminates the kappa locus before lambda gene rearrangement. The KDE rearrangements occur in virtually all Ig-lambda B cells and in one-third of Ig-kappa B cells.

Immunoglobulin Light-Chain Gene Rearrangement in B Cells

Like the Ig heavy-chain gene on chromosome 14, the Ig light-chain genes consist of a series of gene segments in the germline configuration (Fig. 13.16). Two separate genes code for the Ig light chains, the kappa locus on chromosome 2 and the lambda locus on chromosome 22. At the kappa locus, there is a single constant gene segment, 5 joining (J_κ) gene segments, and at least 76 variable (V_κ) gene segments (30 to 35 functional) belonging to 7 sequence-related families.[79,80] In addition, there is a **kappa deleting element (KDE)** located 24 kbp 3′ to the constant region (Fig. 13.17). This element determines deletion of the Igκ constant region in cells producing Ig lambda light chains. The Ig lambda gene locus consists of 52 variable (V_λ) gene segments

(29 to 33 functional) from 10 V_λ families and 7 J_λ (4 to 5 functional) and 7 to 11 C_λ gene segments (4 to 5 functional) occupying 1,140 kb of DNA.[81]

Immunoglobulin light-chain gene rearrangement is similar to that described for the immunoglobulin heavy-chain gene. Selected gene segments are joined together, with loss of the intervening DNA and possible insertion of nucleotides at the junction. The kappa locus rearranges first and then the lambda locus, if necessary. If the lambda locus rearranges, the kappa locus undergoes a secondary recombination through the KDE so that the cell does not produce both types of light chains.

During differentiation of the B cells from precursor stem cells, rearrangement, recombination, and mutation of the immunoglobulin V, D, and J regions ultimately results in functional VJ (light-chain) and VDJ (heavy-chain) genes.

T-Cell Receptor Gene Rearrangement

The T-cell receptor is composed of two of four chains, α, β, γ, and δ, with characteristic structures resembling immunoglobulin V, J, and C regions (Fig. 13.18). The α and δ chains are encoded on chromosome 14 (the δ gene is located inside of the α gene), and γ and β are located on chromosome 7. The four chains form pairs, making two types of receptors, αβ and γδ. T-cell receptor genes have fewer variable gene segments than the immunoglobulin genes, and the genes for the γ and α chains have no diversity regions (Table 13.6). The V regions of the receptor chains undergo gene rearrangement by intrachromosomal recombination as described for the immunoglobulin genes (Fig. 13.18).

Advanced Concepts

Antibody diversity is ensured by three separate events during and after the gene-rearrangement process. The first is the selection of gene segments. The second is the imprecise joining of the segments together with the addition of nucleotide bases at the junction.[82] Third, after the gene-rearrangement process has finished and the B cell

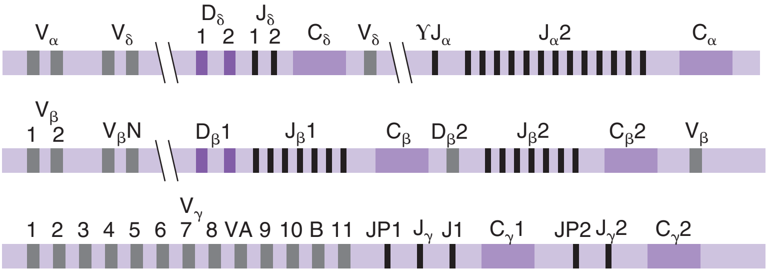


FIGURE 13.18 General structure of the T-cell receptor genes. The gene for the delta T-cell receptor chain is contained in the alpha locus *(top)*. The beta and gamma chains are located at separate loci.

TABLE 13.6 T-Cell Receptor Gene Segments

T-Cell Receptor Chain	Variable Gene Segments*	Diversity Gene Segments	Joining Gene Segments*	Constant Gene Segments
α	54/45	0	60/50	1
β	67/47	2	14/13	2
δ	3	3	4	2
γ	14/6	0	5	2

*Total gene segments/functional gene segments has encountered antigen, somatic hypermutation occurs in the variable regions of the rearranged heavy- and light-chain genes. In addition to the deaminase, the mutation process requires the action of repair systems that further substitute alternate nucleotides in the sequence, resulting in different amino acid substitutions in the antibody protein. Changing of the variable-region sequences underlies the process of **affinity maturation.** As the B cells replicate, those producing antibodies with greater affinity to the antigen are favored, generating subclones of cells that may replace the original reactive clone. Over the course of an infection, therefore, antibodies with increased affinity are produced.

Advanced Concepts

The T-cell receptor δ gene is flanked by TCR δ–deleting elements. Recombination between these elements or between Vα and Jα results in deletion of the TCR δ gene.

Rearrangement of the T-cell receptor chains proceeds in a similar manner as in the immunoglobulin genes. The V, (D), and J segments are joined together with the addition or deletion **(trimming)** of nucleotides at the junctions between the gene segments (Fig. 13.19).

The extracellular domains of the T-cell receptor dimers are held in conformation by interchain disulfide bridges between cysteine residues in the T-cell receptor peptides. The T-cell receptor chains also have a hydrophobic transmembrane region and a short cytoplasmic region. Although most cells express the αβ receptor, λδ receptors can represent a predominant population in certain tissues, such as the intestinal tract.

Detection of Clonality

Gene rearrangements occur independently in each lymphocyte so that a normal population of lymphocytes is **polyclonal (polytypic)** with respect to their rearranged immunoglobulin or T-cell receptor genes. Overrepresentation of a single rearrangement in a specimen cell population can be an indication and characteristic of a lymphoma or leukemia. When over 1% of cells make the same gene rearrangement, the cell population is referred to as **monoclonal (monotypic)** with respect to the rearranged genes.

Advanced Concepts

It is important to distinguish a large monotypic population of tumor cells from a reactive clone or **oligoclone** or reactive clone of cells responding to an antigen. Oligoclones are not only smaller but transient in nature, so they should not be consistently present in serial analyses.

Molecular Analysis of Immunoglobulin Heavy-Chain Gene Clonality

Clonality can be detected by protein and nucleic acid analyses. The first nucleic acid tests for gene rearrangements were performed by Southern blot.[83] With regard to the immunoglobulin heavy-chain gene, restriction sites were mapped in the germline configuration. *Bam*H1, *Eco*R1, and *Hind*III were restriction enzymes commonly used for this procedure (Fig. 13.20). When the germline sequence is rearranged, the restriction sites are moved, created, or deleted, resulting in a unique restriction pattern for every gene rearrangement.

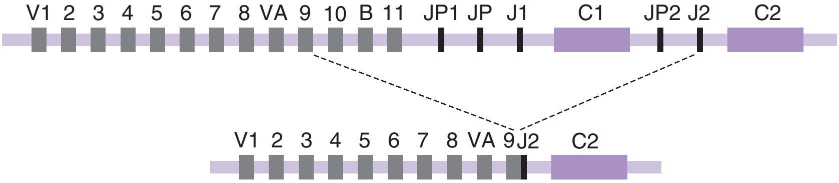

V1 2 3 4 5 6 7 8 VA 9 10 B 11 JP1 JP J1 C1 JP2 J2 C2
V1 2 3 4 5 6 7 8 VA 9 J2 C2

FIGURE 13.19 T-cell receptor gamma gene rearrangement occurs through selection of variable (V) and joining (J) segments.

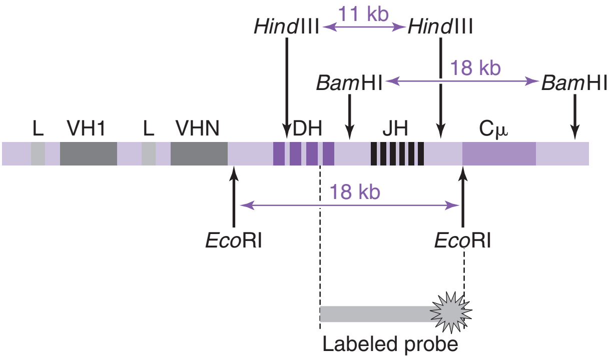


FIGURE 13.20 Restriction map of the germline immunoglobulin heavy-chain gene. The fragments indicated by the arrows will be detectable with the probe shown at the bottom. Gene rearrangement will affect the placement of the restriction sites such that fragments of different sizes will be generated from a rearranged gene.

For the Southern blot, DNA cut with the restriction enzymes is transferred to a nitrocellulose membrane and probed to the joining region of the gene. Normal results should reveal the expected fragments generated from the germline DNA, in the example, an 18-kbp *Eco*R1 fragment, an 18-kbp *Bam*H1 fragment, and an 11-kbp *Hind*III fragment. The normal fragments are visible for two reasons. First, a normal patient specimen contains cells other than lymphocytes that do not undergo gene rearrangement. The second reason for the presence of the germline bands is that only one chromosome in a lymphocyte undergoes gene rearrangement, leaving the homologous chromosome in the germline state. If the first rearrangement fails or is unproductive, then the second chromosome will rearrange. This occurs in fewer than 10% of lymphocytes for the heavy-chain gene.

In a normal specimen, there will be millions of immunoglobulin gene rearrangements, but no one rearrangement is present in high enough amounts to be visible as nongermline bands on the membrane or autoradiogram; thus, only the germline bands will be visible. If, however, 2% to 5% of the cells in the specimen consist of a clone of cells, all with the same gene rearrangement, that clone will be detected by the presence of additional bands different from the germline bands (Fig. 13.21). Interpretation of the results is positive if extra bands are present and negative if only the germline bands are present.

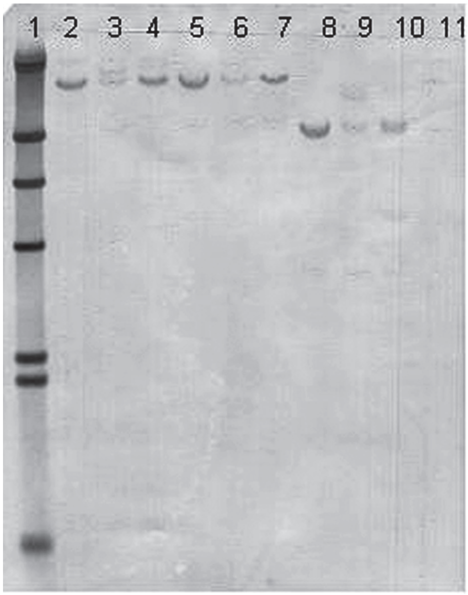


FIGURE 13.21 Colorimetric results from an immunoglobulin heavy-chain gene rearrangement test by Southern blot with colorimetric detection. Lane 1, molecular-weight markers. Lanes 2, 6, and 8 show the normal 18-kbp, 18-kbp, and 11-kbp bands expected from the germline gene in a normal specimen cut with *Eco*R1, *Bam*H1, and *Hind*III, respectively. Lanes 3, 6, and 9 show patient DNA with a monoclonal cell population. Lanes 4, 7, and 10 show a patient with no detectable monoclonality.

Advanced Concepts

Three enzymes are used for this assay to avoid false-positive results due to cross-hybridization artifacts. The chance of true rearranged bands being identical to cross-hybridization patterns for all three enzymes is negligible. In addition, cross-hybridization patterns are constant with the same hybridization conditions, in contrast to true monoclonal gene-rearrangement bands that differ for each monoclonal population.

Analysis of clonality by Southern blot affords the advantage of detecting all gene rearrangements, including incomplete rearrangements involving only the D and J regions of the gene. The Southern blot method is limited by the requirement for at least 20 to 30 µg of high-quality DNA from the specimen. This is not always available, either because the specimen is limiting with respect to cell number or because it is of limited quality, as in paraffin-embedded specimens. Moreover, artifacts such as cross-hybridizations may complicate the interpretation of Southern blot results.

PCR is the most frequent approach for the detection of gene rearrangements. For this method, forward primers complementary to the variable region and reverse primers to the joining or constant regions of the B-cell–rearranged genes are used.[84] The resulting amplicons are resolved by agarose, polyacrylamide, or capillary electrophoresis. A normal cell population will yield amplicons consisting of fragments of different lengths, yielding a cloud or smear of bands, or series of peaks. A monoclonal cell population will yield a predominant or single band or peak, the presence of which is interpreted as a positive result. A limitation of gene rearrangement studies by PCR is the inability to amplify all possible gene rearrangements as primer binding sites are lost during recombination and somatic mutation.

A limitation to the detection of clonality by PCR is the loss of the FR3 primer-binding site due to the gene-rearrangement process, which may destroy or remove the sequences bound by the FR3 primer. To address this issue, some laboratories use additional forward primers that bind to framework 2 (FR2) and framework 1 (FR1). Other methods utilize primers to the leader region in addition to the FR primers.[84,85] These primers increase the number of gene rearrangements that can be amplified and therefore detected by this assay. Deletion of the IgH locus can occur in some lymphomas, precluding amplification with any primers.[86] If the gene rearrangement cannot be amplified, cytogenetics may be used to assess clonality at the immunoglobulin heavy-chain gene locus.[87]

For immunoglobulin heavy-chain gene rearrangement, different approaches to the variable-region primer-binding sites have been taken (Fig. 13.22). The immunoglobulin heavy-chain gene-variable region is divided into two types of domains. The **complementarity-determining regions (CDRs)** code for the amino acids that will contact the antigen. The CDRs, therefore, are the most variable, or unstable, in sequence. The **framework regions (FRs)** code for the amino acids that have more of a structural role in the antibody protein and are more stable in sequence. Standard methods for clonality utilize a forward **consensus primer** to the innermost framework region (FR3) and the reverse primer complementary to the joining region. The consensus primer has sequences that match the most frequently occurring sequences in the FR3 region and may not be identical to any one sequence. Enough nucleotides will hydrogen bond, however, to match most rearranged variable regions. Primers directed at the diversity region are useful for amplification of the germline configuration of the immunoglobulin heavy-chain genes as well as

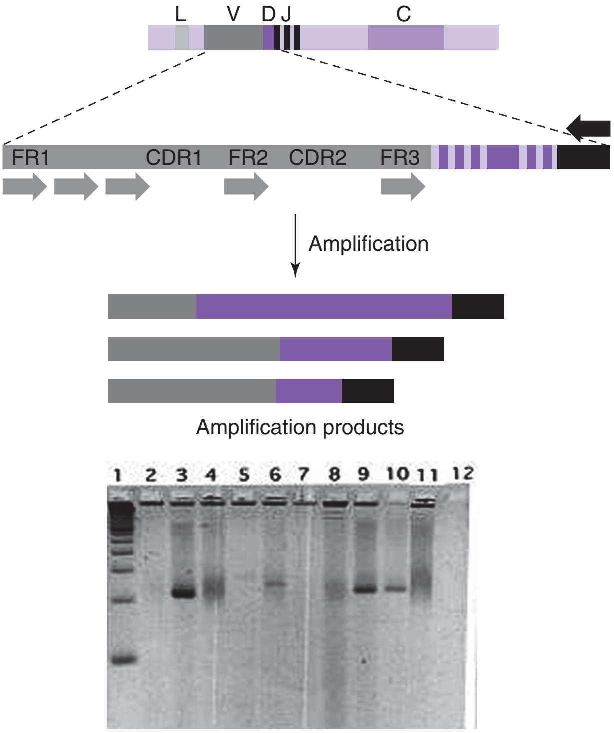


FIGURE 13.22 Immunoglobulin heavy-chain gene rearrangement by PCR with amplification from the variable region. Forward primers complementary to the variable region and reverse primers complementary to the joining region are used to amplify the diversity region *(top)*. In a polyclonal specimen, amplification products in a range of sizes will result. These products produce a dispersed pattern on an ethidium bromide–stained agarose gel (lanes 4, 8, and 11). If at least 1% of the sample is representative of a monoclonal gene rearrangement, that product will be amplified preferentially and revealed as a sharp band by gel electrophoresis (lanes 3, 6, 9, and 10).

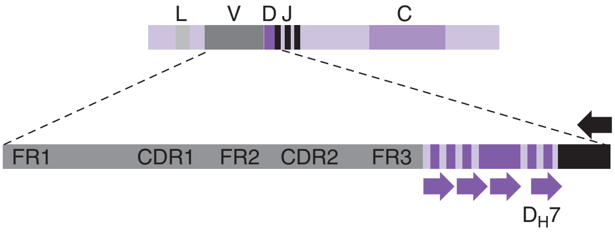


FIGURE 13.23 Immunoglobulin heavy-chain gene rearrangement by PCR with amplification from the diversity region. Forward primers complementary to the diversity region and reverse primers complementary to the joining region yield a polyclonal pattern in normal samples. The D_H7 primer will yield a specific 350-bp product from the unrearranged (germline) gene due to the short distance between the D_H7 gene segment and the joining-region primer.

D_H–J_H rearrangements.[88] Primers complementary to the seven family-specific sequences of the diversity region and a primer to the 5′-most joining region (J_H1) are used to amplify the D_H–J_H junction. The primer complementary to the D_H7-27 segment, the sequences closest to the joining region, will yield a defined product from the Ig heavy-chain gene in the germline configuration (Fig. 13.23). Diversity-region primers are also useful for targeting incomplete rearrangements where the variable-region primer-binding sites are lost.[89,90]

Another approach to the detection of B-cell clonality is to make patient-specific primers. For this method, a consensus primer is used to amplify the rearranged gene from a positive specimen. The amplification product is then purified from the gel or from the PCR reaction mix and sequenced. Primers exactly matching the variable-region sequence of that specimen are then manufactured for use on subsequent samples. The advantage of this method is that it is more sensitive because fewer or none of the gene rearrangements from normal cells are amplified. Furthermore, the tumor load can be measured quantitatively by real-time PCR (RT-PCR). The disadvantage of this approach is that the method is more time-consuming to perform, and the primers used are patient specific. Moreover, in a patient with a chronic condition, for whom monitoring is likely to be done, tumor cells may undergo further mutation in the variable region and inactivate the specific primer binding, requiring the manufacture of new primers.

Immunoglobulin Light-Chain Gene Rearrangements

The immunoglobulin light-chain genes are also targets for clonality detection. In addition to protein analysis, Southern blot, PCR and RT-PCR have been used to detect light-chain gene clonality.[91–93] Targeting the immunoglobulin light-chain genes is especially useful for tumors arising from terminally differentiated B cells (plasma cells) that have undergone extensive somatic hypermutation at the heavy-chain gene locus. In these tumors, the rearranged heavy-chain genes are frequently unamplifiable, yielding false-negative results because of accumulation of base changes in the variable-region primer-binding sites. Most of these tumors are amplifiable, however, at the light-chain genes.[94]

Gene rearrangements in the kappa light-chain locus are amplified using primers complementary to the sequence families of the V_κ region or to the intron between the J_κ regions and C_κ. Opposing primers are complementary to C_κ and to one or more of the five joining regions. Alternately, the KDE can be used for a primer-binding site (Fig. 13.24). Using the KDE allows detection of lambda-expressing cells that have deleted J_κ or C_κ.[92]

Immunoglobulin heavy-chain gene rearrangement precedes Ig light-chain gene rearrangement during early B-cell differentiation in the bone marrow. Following successful IGH recombination, the Ig light-chain loci then rearrange, first at the Ig kappa locus. If the kappa rearrangement is nonfunctional, the Ig lambda locus rearranges. With functional heavy- and light-chain gene rearrangements, the cell will develop into a mature naïve B cell. Therefore, detection of clonality at the Ig_λ locus on chromosome 22 is also useful for confirming or monitoring diagnosis of B-cell leukemias and lymphomas. Forward primers complementary to V_λ gene segments and reverse primers to the J_λ and C_λ gene segments are frequently used for these assays (Fig. 13.25).

T-Cell Receptor Gene Rearrangements

Like B-cell clonality and B-cell malignancies, clonality of T-cell receptor gene rearrangements may demonstrate that a high white cell count is of T-cell origin or the presence of a clonal T-cell infiltrate accompanying another malignancy. Clonality may also be used to monitor treatment efficacy. Clonal T-cell receptor gene

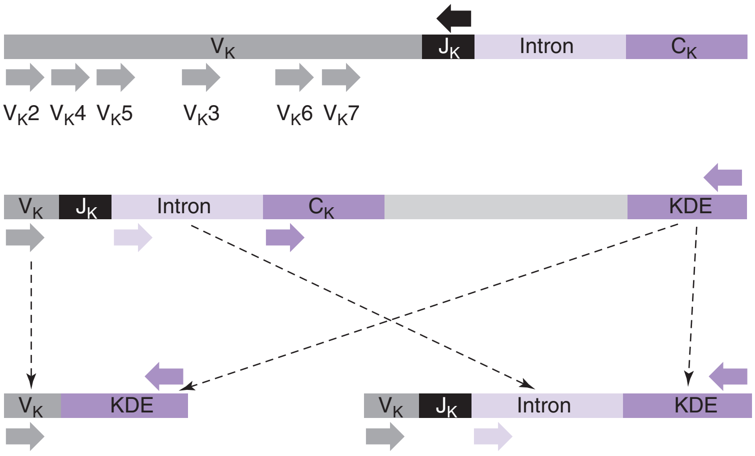


FIGURE 13.24 The immunoglobulin light-chain kappa locus can be amplified by standard PCR procedures from the variable region, the intronic region, or the kappa-deleting element.

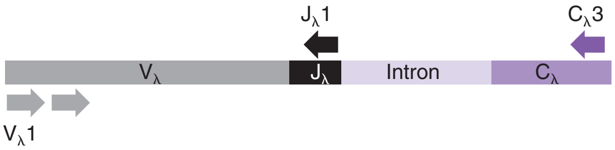


FIGURE 13.25 The immunoglobulin light-chain lambda locus can be amplified by standard PCR procedures from the variable region to the joining region or the constant region. Reverse transcriptase PCR is used with the constant-region primers to eliminate the intron between the constant and joining regions.

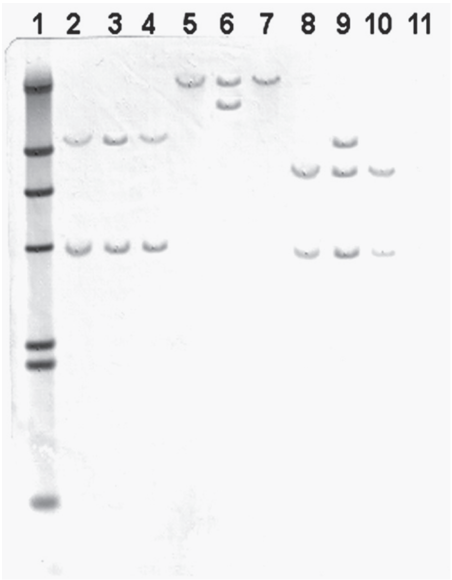


FIGURE 13.26 T-cell receptor gene rearrangements detected by Southern blot using a probe mixture of sequences complementary to $J_{\beta}1$ and $J_{\beta}2$. Lane 1, molecular-weight markers. Lanes 2, 5, and 8, normal control cut with *Eco*R1, *Bam*H1, and *Hind*III, respectively. Lanes 3, 6, and 9, a negative specimen cut with *Eco*R1, *Bam*H1, and *Hind*III, respectively. Lanes 4, 7, and 10, positive specimen cut with *Eco*R1, *Bam*H1, and *Hind*III, respectively. Note the additional bands in the positive specimen lanes.

rearrangements are detected in a manner similar to the immunoglobulin gene rearrangements by Southern blot, PCR, and sequencing.[95] For Southern blot studies, probes complementary to the variable, joining, and diversity regions of the T-cell receptor genes are used to detect monoclonal populations. Just as with immunoglobulin gene rearrangements, no one gene rearrangement should be visible in a normal specimen. The presence of a large monoclonal population is revealed by the bands detectable in addition to the germline bands seen in the normal control (Fig. 13.26).

The T-cell receptor gene rearrangement assays done by PCR most often target the *TCRγ* gene. Assays are also performed on the *TCRβ* and *TCRδ* genes. Detection of gene rearrangements in *TCRα* is difficult due to the 85-kb length of the J_{α} gene segments. *TCRα* gene rearrangements may be inferred from *TCRδ* gene deletions.[96] Primers are designed complementary to the rearranged gene segments (Fig. 13.27). Multiple primer sets are often used to ensure detection of the maximum potential gene rearrangement.[97] PCR will yield a product consistent with a single-gene rearrangement in a positive sample, whereas a normal sample will yield a polyclonal pattern (Fig. 13.28). Due to the limited range of lengths of the population of rearranged T-cell receptor genes, heteroduplex analysis may be used to improve the resolution of the polyclonal and monoclonal patterns.

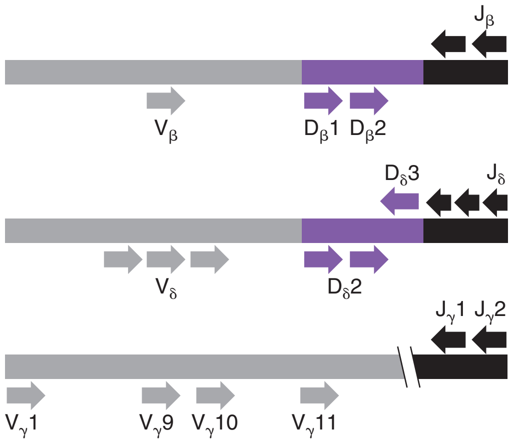


FIGURE 13.27 T-cell receptor gene rearrangements detected by PCR use primers to the variable, diversity, and joining regions of the rearranged genes.

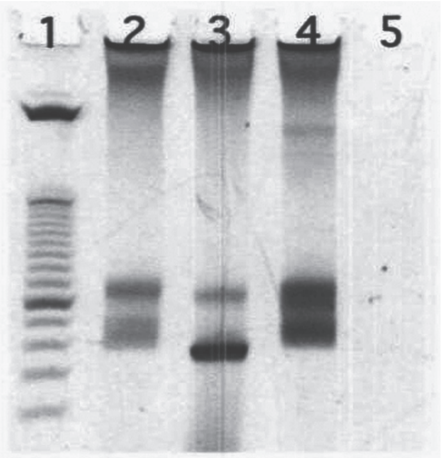


FIGURE 13.28 T-cell receptor gene rearrangement by PCR and heteroduplex analysis. Bands were separated by polyacrylamide gel electrophoresis and stained with ethidium bromide. Lanes 2 and 4 show a polyclonal pattern. A positive result (monoclonal pattern) is shown in lane 3. Lane 1, molecular-weight marker; lane 5 reagent blank.

In 2003, a primer set was designed by the European Commission Biomed-2 group to amplify an increased number of possible immunoglobulin heavy-chain gene rearrangements.[84] This set comprises 107 primers multiplexed in 18 PCR reactions. In addition to immunoglobulin heavy-chain, T-cell receptor gene rearrangements, these primers are directed at two translocations, t(14;18) and t(11;14) (see following discussion).

Advanced Concepts

T-cell receptors do not undergo somatic hypermutation when a given T cell is exposed to antigen. Therefore, the specificity of a given T-cell clone maintains a common antigen specificity, which is the **clonotype** of that population. Massive parallel sequencing of T-cell receptors and alignment of the data to the germline gene sequences can reveal a rearrangement. By taking a tally of the reads, a full picture of the T-cell receptor clonotype can be generated.[98]

Banding Patterns

Interpretation of clonality by PCR depends on gel banding patterns. False-negative results may arise from primers that do not match the gene rearrangement in the tumor. On the other hand, artifactual single bands will produce false-positive results. Patterns of multiple single bands may arise from specimens with low cell numbers, such as cerebrospinal fluid or paraffin sections. In this case, usually more than one or two single bands will occur. The bands may not indicate monoclonality but, rather, that only a few cells are present. These cells may or may not be malignant. False single-band patterns may also occur within polyclonal smears detected on nondenaturing polyacrylamide gels (PAGEs). Heteroduplex analysis is recommended if amplicons are resolved by nondenaturing PAGEs.

Similar problems occur with resolution by capillary electrophoresis where peak heights and widths are compared to distinguish wide polyclonal peaks from narrow monoclonal spikes.[99] Pretreatments such as eosin staining may yield false single bands by capillary electrophoresis.[100]

Mutations in Hematological Malignancies

Lymphoid Malignancies

Immunoglobulin and T-cell receptor gene rearrangements are normal and take place regardless of the presence of malignancy. The unique nature of the gene-rearrangement system with cell-specific antibody and antibody receptor formation is exploited in finding abnormal cell populations clonally derived from these cell-specific features. Targeting these gene rearrangements, however, can only be applied to tumors arising from lymphocytes. Other types of malignancies, such as myeloid tumors, cannot be analyzed using these targets.

Often, the development of the cancer state results in DNA anomalies, such as translocations or other types of mutations. These can be applied to the clinical discovery and monitoring of tumors. These abnormal genetic events aid in the diagnosis of specific types of hematological tumors, such as the translocations that are highly associated with chronic myelogenous leukemia, promyelocytic leukemia, or follicular lymphoma. Translocations are the exchange of DNA between chromosomes. If the translocation disrupts the activity or expression of oncogenes or tumor-suppressor genes, a cancer phenotype can occur. Several translocations are frequently found in certain types of tumors.

t(14;18)(q32;q21)

The reciprocal translocation between the long arms of chromosomes 14 and 18 moves and dysregulates the *BCL2* gene located on chromosome 18q21.3. *BCL2* (B-cell leukemia and lymphoma 2 or B-cell CLL/lymphoma 2) is an oncogene. The gene product of *BCL2* is a member of a group of related proteins that control apoptosis (cell death initiated by internal cellular signals). The Bcl2 protein inhibits apoptosis in B lymphocytes, that is, enhances survival of cells that normally would die. Survival of genetically damaged cells may contribute to the development of tumors. One of the most frequent hematological malignancies, follicular lymphoma, is associated with the t(14;18) translocation.

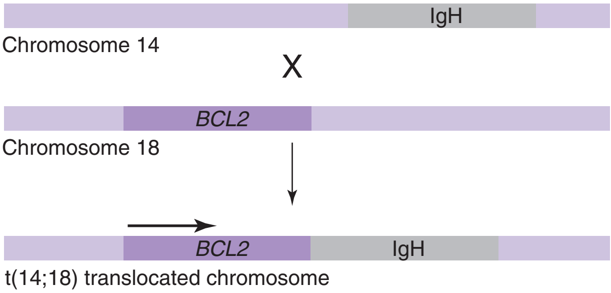


FIGURE 13.29 The t(14;18) translocation moves the *BCL2* gene intact to the long arm of chromosome 14 next to the joining region of the immunoglobulin heavy-chain gene (IgH). The translocation breakpoints on chromosome 18 are 3 + to the *BCL2* gene (arrow indicating the direction of transcription).

The translocation partner of chromosome 18 is chromosome 14 in the vicinity of the immunoglobulin heavy-chain gene. The translocation may occur through cryptic recognition sites for the gene-rearrangement recombinase enzymes on chromosome 18. As a result, an abnormal interchromosomal exchange occurs instead of the normal intrachromosomal exchange events described in the previous section (Fig. 13.29). There are several breakpoints in the chromosome 18 region 3′ to the *BCL2* gene, most of which occur in the major breakpoint region (MBR). About 10% to 20% of the breakpoints fall into a cluster closer to the *BCL2* gene, thousands of bases from the MBR, in the minor cluster region (MCR). An intermediate cluster region (ICR) and other breakpoints outside of MBR and MCR have also been reported.[101] Breakpoints in this and other translocations are associated with double-helix disruption or fragile sites in the DNA.[102]

Molecular detection of the t(14;18) translocation is performed by a variety of methods. Southern blot with a probe to the MBR region of chromosome 18 will reveal the translocation by the presence of bands different from those expected from a normal chromosome 18 (Fig. 13.30). The translocation is easily and rapidly detected by PCR or qPCR. Forward primers to chromosome 18 and reverse primers complementary to the Ig heavy-chain joining region will yield a product only if the two chromosomes have been joined by the translocation. The amplicons are visualized by gel electrophoresis, as shown in Figure 13.31, or with a qPCR probe. The t(14;18) translocation can be used to quantify tumor load by qPCR. Several methods are available for this analysis, and although laboratory methods differ, the results

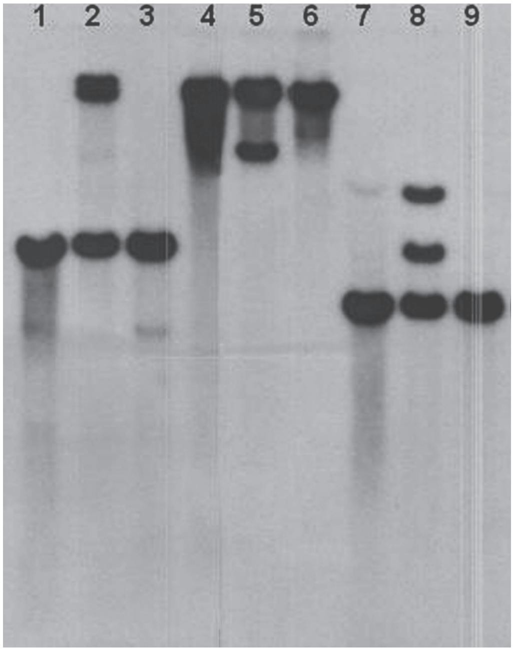


FIGURE 13.30 Analysis of the t(14;18) gene translocation by Southern blot with chemiluminescent detection. Lanes 3, 6, and 9 are the normal control cut with *Eco*R1, *Bam*H1, and *Hind*III, respectively. Lanes 1, 4, and 7 show results from a normal patient specimen. Lanes 2, 5, and 8 are results from a specimen positive for the t(14;18) translocation. Note the extra bands in the positive specimen.

were found to be reasonably consistent in comparison testing.[103] Most methods include a standard curve for regression analysis of the test sample measurements (Fig. 13.32). Alternatively, internal controls, such as known amounts of plasmid DNA, may be added to the test specimens. In this method, the amount of translocated cells (or translocated chromosomes) is determined relative to the internal control. There are several approaches to reporting the final results. Most frequently, the results are reported as the percentage of translocated cells in the specimen tested. For instance, the raw number of translocated cells is determined using the standard curve of threshold cycles versus the known cell count. The raw number of cells is then divided by the number of total cells represented in the PCR reaction.

For example, for a dilution series ranging from 50 to 10,000 translocated cells in 1 million untranslocated cells, the standard curve generates the formula $y = 1.5631[\mathrm{Ln}(x)] + 42.396$, where y is the threshold cycle number, and x is the number of translocated cells (50 ng of DNA was used per sample). Using this formula, if a given sample crosses the fluorescence threshold at $y = 36$ cycles (average of duplicate measurements), the number of cells (x) is approximately 60 cells. Assuming that 50 ng of DNA represents 7,500 cells (1 ng of DNA = approximately 150 cells), $60/7{,}500 = 0.008$. $0.008 \times 100 = 0.8\%$ translocated cells in the specimen. For the t(14;18) translocation, sensitivities of 0.0025% have been reported, with a linear range of 0.01% to 10%.[104]

The main limitation of any PCR procedure targeting the t(14;18) translocation is the inability of the primers to detect all of the possible breakpoints on chromosome 18. Primer pairs, sets of primer pairs, and novel probe methods have been designed to address this problem.[86,105] Test reports should include an estimation of false-negative results expected due to breakpoints that remove the binding sites for the primers used. Unless a translocation has been previously observed by a given PCR method, a negative result should acknowledge that a translocation may be present but undetectable with the primers used.

t(11;14)(q13;q32)

The t(11;14)(q13;q32) translocation joins the immunoglobulin heavy-chain gene region on chromosome 14 with part of the long arm of chromosome 11. The cyclin D1 *(CCND1)* gene, also called the parathyroid adenomatosis 1 gene *(PRAD1)* or *BCL1,* on chromosome 11 is attached to the long arm of chromosome 14 in the intron between the immunoglobulin heavy-chain gene joining and constant regions. The translocation increases expression of *CCND1,* resulting in passage of the cell cycle from the G1 to the S phase of the cell cycle. This translocation is found primarily in mantle cell lymphoma (MCL) but may also be present in chronic lymphocytic leukemia, B-prolymphocytic leukemia, plasma cell leukemia, multiple myeloma, and splenic lymphoma. The t(11;14) translocation is thought to be a definite characteristic of MCL; however, only 50% to 70% of MCLs have a detectable t(11;14).

Mantle cell lymphoma without the t(11;14) translocation is likely to show expression of the transcription

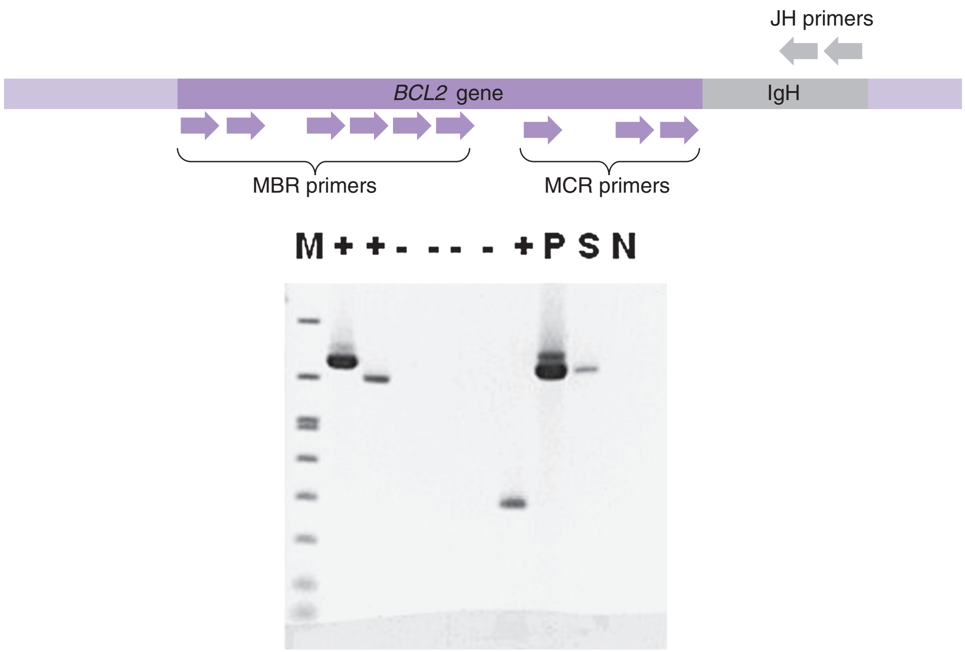


FIGURE 13.31 Analysis of the t(14;18) gene translocation by PCR with agarose gel electrophoresis. Several primers are used for detection of the translocated chromosome *(top)*. Because the forward primers are on chromosome 18 and the reverse primers are on chromosome 14, a PCR product will be generated only from the t(14;18) translocated chromosome *(bottom)*. *M,* molecular-weight marker; *+,* positive specimen; *-,* negative specimen; *P,* positive control; *S,* sensitivity control; *N,* negative control.

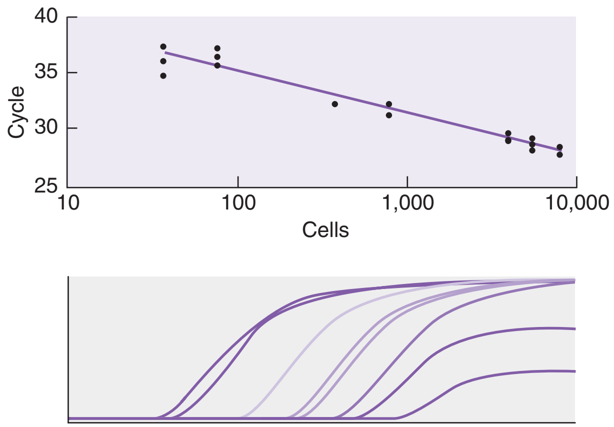


FIGURE 13.32 The t(14;18) translocation analysis by qPCR. A standard curve is established using cultured cells with the t(14;18) translocation counted and mixed at different proportions with cultured cells that do not have the translocation. The illustration represents DNA isolated from each mixture of cells analyzed by qPCR with a TaqMan probe.

factor SOX11. SOX11 therefore represents an important performed by IHC or western blot.[106]

Methods for detection of t(11;14) are similar to those used for t(14;18) translocation described previously. Southern blot methods have been replaced mostly with PCR and RT-PCR,[107] but due to the variation in breakpoints, primers can miss some translocations. In these cases, FISH or flow-cytometry analysis may be preferable.[108] Of the breakpoints on chromosome 11, 80% are in the major translocation cluster 5′ to the *CCND1* gene (Fig. 13.33). The rest are dispersed in other areas 5′ or 3′ of the gene. PCR analysis detects 40% to 60% of the translocation breakpoints.[109] A PCR product will result only if the translocation has occurred (and the primer binding sites are maintained). The PCR product can be detected by qPCR fluorescence or by gel or capillary electrophoresis.

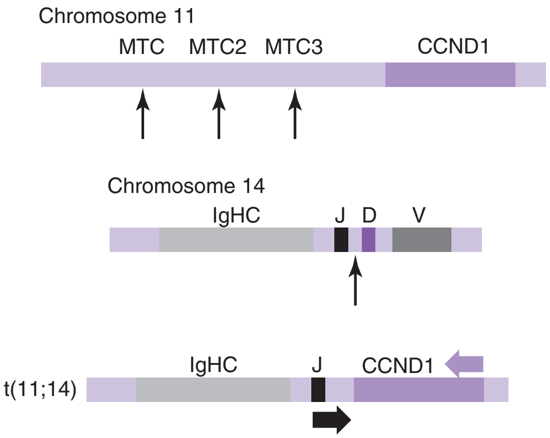


FIGURE 13.33 The t(11;14) breakpoints in chromosome 11 *(top)* and chromosome 14 *(center)* are indicated by the vertical arrows. Primers are complementary to the joining region of the immunoglobulin heavy-chain gene and the *CCND1* gene *(bottom)*.

Advanced Concepts

A variety of molecular events can result in the overexpression of *c-myc*. The DNA virus Epstein–Barr virus (EBV) is associated with Burkitt lymphoma. EBV induces cell proliferation and, thus, provides more opportunity for translocation events. Certain types of another DNA virus, human papillomavirus, inserted into the vicinity of *c-myc* cause overexpression of the gene.[110]

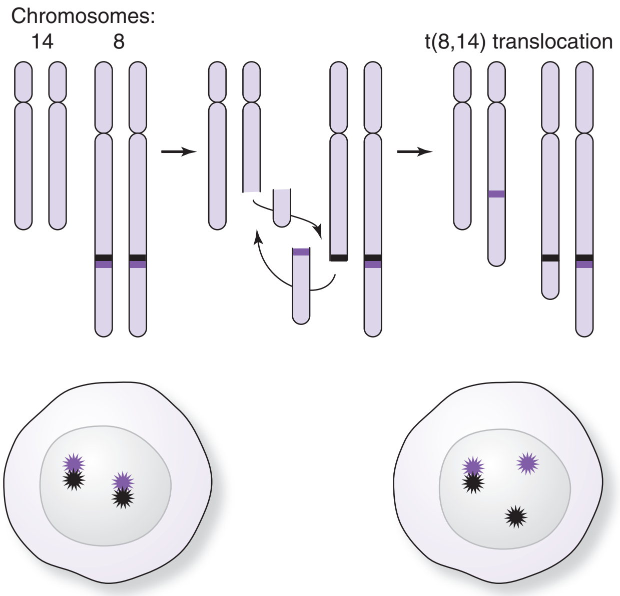


FIGURE 13.34 The t(8:14) breakpoint detected by CISH (Invitrogen). Two probes labeled with biotin or digoxigenin are complementary to sequences flanking the chromosome 14 breakpoint. In the absence of the translocation *(left)*, the probes will appear next to one another in the nucleus. The translocation *(right)* will move one probe to chromosome 14, leaving the other behind on chromosome 8, resulting in separation of the signals in the nucleus.

t(8;14)(q24;q11)

The avian myelocytomatosis viral oncogene homolog *(c-myc)* gene on chromosome 8 is one member of a gene family including *n-myc* and *l-myc*. The *c-myc* gene codes for a helix-loop-helix/leucine zipper transcription factor that binds to another protein, Max, and activates transcription of other genes. The t(8;14) is associated with Burkitt lymphoma; in addition, translocations at (2;8) and t(8;22) are found in about 10% of Burkitt lymphomas.

In the t(8;14) translocation, the breakpoints on chromosome 8 are spread over a 190-kbp region 5′ to and within the *c-myc* gene. As a result of the t(8;14) translocation, the *c-myc* gene is separated from its normal promoter and regulatory region and moved into the switch recombination region of the immunoglobulin heavy-chain gene on chromosome 14. In the t(2;8) and the t(8;22) translocations, the chromosome 8 breakpoints are 3′ to the gene, and *c-myc* is moved into the immunoglobulin kappa or lambda locus, respectively. Translocations of *c-myc* into the T-cell receptor alpha gene also occur.[111]

In the laboratory, *c-myc* translocations were commonly detected by Southern blot analysis with a probe complementary to exon 3 of the *c-myc* gene; for instance, a ^{32}P- or digoxigenin-labeled 1.4-kb *Cla*I-*Eco*RI restriction fragment. Interphase FISH and CISH are now used to detect the t(8;14) translocation with separate probes to chromosomes 8 and 14 or with chromosome 14 probe pairs (Fig. 13.34). Amplification of the gene can be detected by FISH using a 120-kb *c-myc* (8q24.12-q24.13) fluorescent probe. Myc protein expression is measured by IHC.[112]

Myeloid Malignancies

t(9;22)(q34;q11)

The t(9;22) translocation is a reciprocal exchange between the long arms of chromosomes 9 and 22. The translocation generates the Philadelphia chromosome (Ph1), which is present in 95% of cases of chronic myelogenous leukemia (CML), 25% to 30% of adult acute lymphoblastic leukemia (ALL), and 2% to 10% of pediatric ALL. The breakpoints of the t(9;22) translocation occur within two genes, the breakpoint cluster region *(BCR)* gene on chromosome 22 and the cellular counterpart of the Abelson leukemia virus tyrosine kinase (c-*abl*) on chromosome 9 (Fig. 13.35). The result of the translocation is a chimeric or **fusion gene** with the head of the *BCR* gene and the tail of the c-*abl* gene. Both genes are tyrosine kinases; that is, they phosphorylate other genes at tyrosine residues. The fusion gene is also a kinase but has aberrant kinase activity. There are two major forms of the *BCR/abl* fusion gene, joining either exon 13 or 14 (b2 or b3) of the *BCR* gene to c-*abl* exon 2 (a2; Figure 13.36). The b2a2 or b3a2 fusion genes code for a 210-kd protein, p210. A third form of the fusion gene joins exon 1 of *BCR* with exon a2 of c-*abl,* resulting in the expression of an e1a2 transcript, which codes for a p190 protein. Another, less common, fusion junction occurs at exon 19 of the *BCR* gene (c3). The c3a2 transcript encodes a p230 protein. All of the fusion proteins have been observed in CML; however, p190 occurs mostly in ALL. The common rationale for the detection of translocations by PCR is used for t(9;22); that is, forward primers are designed to hybridize to the *BCR* gene on chromosome 22 and reverse primers to chromosome 9 in the c-*abl* gene (Fig. 13.37). A product will result only if the two genes are joined by the translocation. Due to the length of the introns that separate the primer-binding sites, nested RT-PCR was initially utilized.[113] To minimize the risk of false-positive results, this method may require confirmatory testing, especially for adult ALL.[114] For optimal performance, primers capable of detecting both major and minor breakpoints in the *BCR* gene are utilized. An internal **RNA integrity control** (or amplification control) is included for each sample to avoid false-negative results from poor RNA quality or inadequate cDNA synthesis. Transcripts from the *abl* or *BCR* genes are used most frequently for the RNA integrity control for the t(9;22) translocation, but other unique genes with constitutive expression have been used. Target amplicons are detected by agarose gel electrophoresis with ethidium bromide staining (Fig. 13.38) or by capillary gel electrophoresis. Fluorescent dye–labeled primers are required for the latter detection method.

Quantitative PCR provides an estimation of treatment response, especially with targeted therapies for CML and ALL. Although cytogenetic methods, especially FISH and standard RT-PCR, are most practical for diagnosis and in the early stages of treatment, quantitative PCR provides a valuable estimation of tumor load over the course of treatment.[115–117]

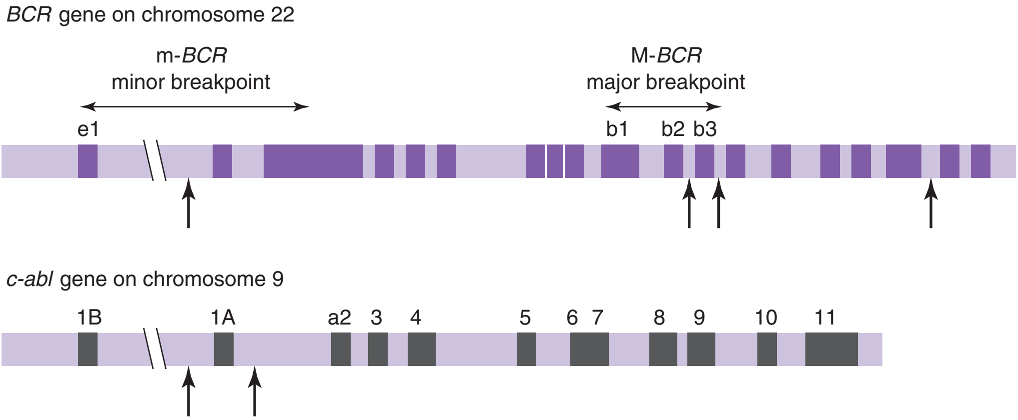


FIGURE 13.35 The t(9;22) translocation begins with breakage of chromosomes 9 and 22 in introns of the *BCR* and c-*abl* genes (arrows).

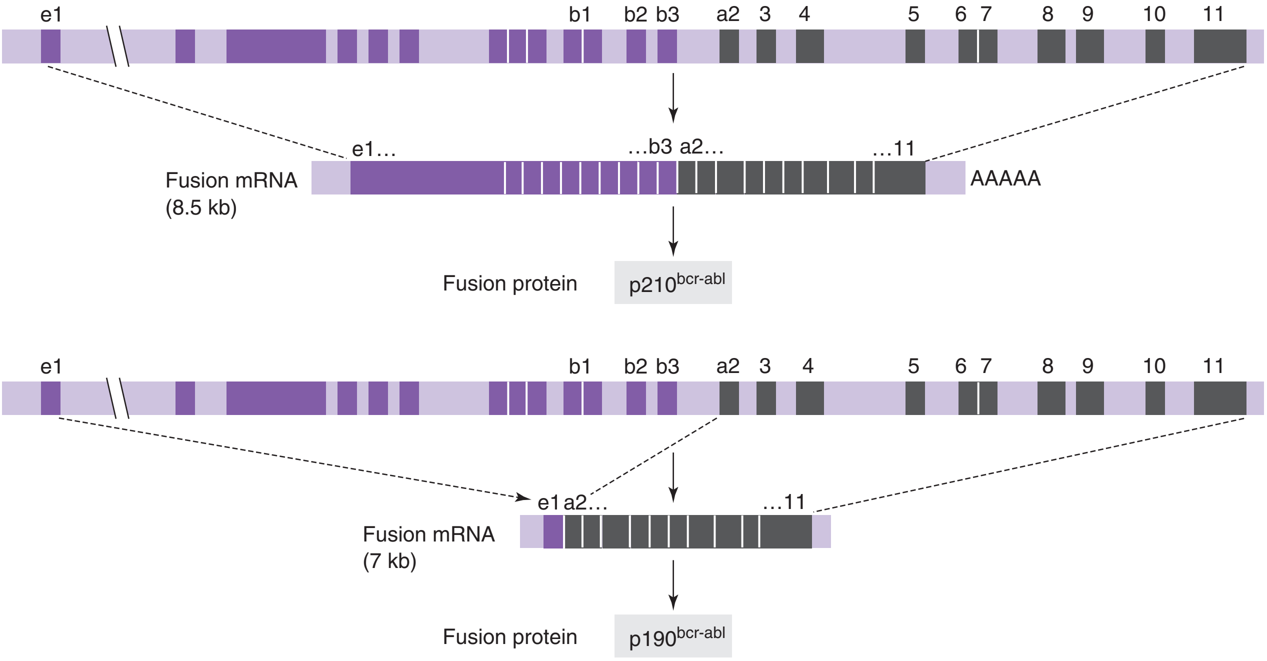

e1
b1
b2 b3
a2
3
4
5
6 7
8
9
10
11
e1…
…b3 a2…
…11
Fusion mRNA
(8.5 kb)
AAAAA
Fusion protein
p210bcr-abl
e1 a2…
…11
Fusion mRNA
(7 kb)
Fusion protein
p190bcr-abl

FIGURE 13.36 p210 and p190 are the two main fusion proteins produced by the t(9;22) translocation. They differ in the amount of the *BCR* gene that is attached to the c-*abl* gene.

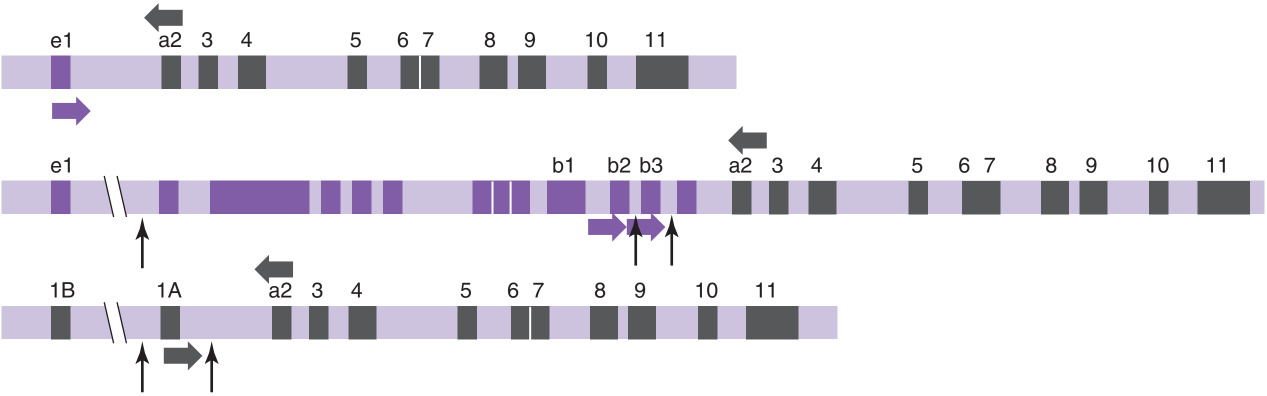

e1
a2
3
4
5
6 7
8
9
10
11
e1
b1
b2 b3
a2
3
4
5
6 7
8
9
10
11
1B
1A
a2
3
4
5
6 7
8
9
10
11

FIGURE 13.37 Location of primers (horizontal arrows) for PCR analysis of the minor breakpoint region *(top)* and the major breakpoint region *(middle)*. The c-*abl* gene itself may be used as an amplification control *(bottom)*. The intron and exon lengths are not drawn to scale. Vertical arrows denote locations of breakpoints.

Because transcript levels are being measured using RT-PCR, it is important to stabilize the specimen RNA on receipt, for example, by resuspending the white blood cells in protective buffers. Another recommendation is to collect sufficient peripheral blood for analysis to avoid false-negative results.[118] The primers used for this method are similar to those used for standard PCR. A fluorescent probe provides the signal. A standard curve or a high positive, low positive (sensitivity) control and negative control should accompany each run. Frequently used methods reported measurements as a ratio of the *BCR-abl* transcript level to the RNA integrity control,

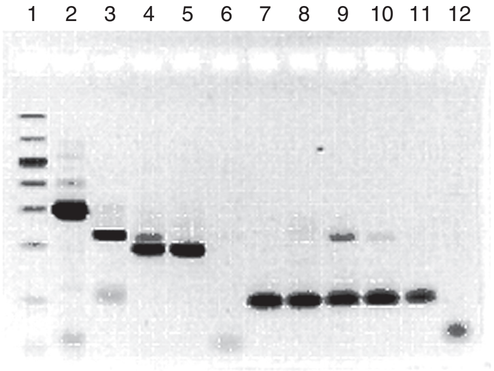


FIGURE 13.38 Results of a standard RT-PCR test for the t(9;22) translocation. Lane 1, molecular-weight marker; lane 2, b3a2 breakpoints; lane 3, b2a2 breakpoints; lanes 4 and 5, e1a2 breakpoints; lane 6, negative specimen; lanes 7 to 11 are RNA integrity (amplification) controls for specimens in lanes 2 to 6; lane 12, reagent blank.

usually the *abl* transcript, the *BCR* transcript, or the transcript of a housekeeping gene, such as *G6PDH*.[115,119] For example, a standard curve for transcript number generates the formula $y = -1.7318[Ln(x)] + 48.627$, where y is the threshold cycle number, and x is the number (or dilution) of transcripts. If quantitative PCR analysis of the patient specimen RNA yielded a threshold cycle number of 39 (average of duplicate samples) for *BCR-abl* transcripts and a threshold cycle number of 30 for the *abl* transcripts, then solving for x yields 300 *BCR-abl* transcripts and 50,000 *abl* transcripts in the sample. Thus, (300/50,000) × 100 = 0.6%.

Data are still being collected with regard to the clinical significance of the quantitative results. A three-log drop in transcript levels and a *BCR-abl/abl* × 100 level below 0.05% have been proposed as indicators of good prognosis.[120,121]

Advanced Concepts

A slightly different formula, [*BCR-abl*/(*abl* – *BCR-abl*)] × 100, for the transcript ratio is used if the *abl* primers also amplify the *BCR-abl* translocation.[122] In the calculation shown in the text, *BCR-abl/abl* × 100 yields 0.6%; [*BCR-abl*/(*abl* – *BCR-abl*)] × 100 yields 0.85%.

Therapeutic tyrosine-kinase inhibitors that target the chimeric BCR-abl protein have changed CML from a deadly disease to a chronic condition, treatable by a pill a day. With the success of these agents, and to detect resistances, efforts have been made to establish a standardized measurement scale to provide consistent results among laboratories.[123,124] Reference standards were established to be used to normalize measurements from diverse laboratory methods.[125] These reference standards were protected RNA molecules or "armored RNA." Armored RNA is an engineered MS2 phage protein complex encapsulating an RNA fragment of the *BCR-abl* target gene designed for use as a qPCR calibration standard. The protected RNA is resistant to RNase digestion and shows high stability in plasma and other fluids. Sets of known concentrations of these reference standards are distributed to laboratories, where they are included in the laboratory measurement. The results are sent to a test center, where the testing laboratory results are adjusted to the actual numbers for the standards with a correction parameter (CP) or conversion factor (CF). The CP or CF is then applied to standardize the *BCR-abl* test results.[126] Alternatively, commercial reagent sets containing internal calibrators and controls may be used. These systems will generate results on the International Scale without the requirement for adjustments.

t(15;17)(q22;q11.2-q12)

The t(15;17)(q22;q11.2-q12) reciprocal translocation between the long arms of chromosomes 15 and 17 results in fusion of the retinoic acid receptor alpha *(RARA)* gene on chromosome 15 with the myelocytic leukemia (*MYL* or *PML*) gene on chromosome 17. Both genes contain zinc finger–binding motifs and therefore bind DNA as transcription factors. The *PML/RARA* fusion is found specifically in promyelocytic leukemia. The presence of this translocation is also a predictor of the response to retinoic acid therapy that is used as a treatment for this disease. The translocation forms a fusion gene with the first three (type A or S translocation) or six (type B or L translocation) exons of the *PML* gene joining to exons 2 to 6 of the *RARA* gene (Fig. 13.39).

Test methods similar to those described previously for *BCR-abl* translocation are used to detect t(15;17), although there is no current standardization other than standard curves generated from reference standards.

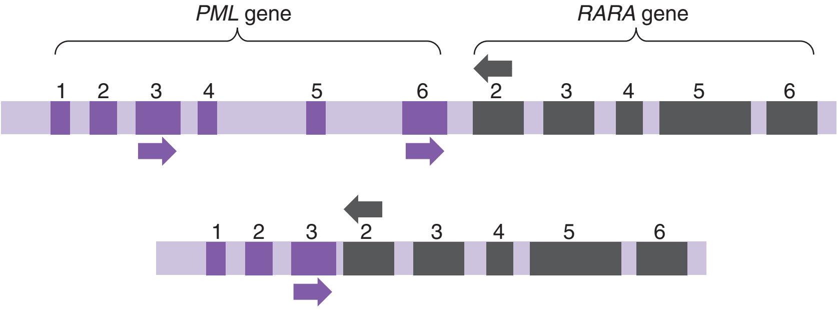


FIGURE 13.39 Location of primers for PCR analysis of the long *(top)* and the short *(bottom)* translocated genes. The first three or six exons of the *PML* gene are fused to exon 2 of the *RARA* gene. The intron and exon lengths are not drawn to scale.

Reverse transcriptase PCR and RT-qPCR are most frequently used. Primers complementary to sequences in exon 3 or 6 of the *PML* gene and exon 2 of *RARA* generate products only if the translocation has occurred. The presence of the translocation product is interpreted as a positive result. As with any test of this type, an amplification control is required to avoid false-negative results. For quantitative PCR, results normalized to an internal control, using calculations as described previously, yield the most consistent day-to-day results (lowest coefficient of variance).[127]

Advanced Concepts

In the t(15;17) translocation, the reciprocal fusion gene *RARA/PML* is also expressed in 70% to 80% of cases of APL. Nonreciprocal events can produce either fusion alone as well.[128] Although the reciprocal fusion may participate in the tumorigenesis process, it does not predict response to all-trans retinoic acid therapy.[129] The presence of the reverse transcript may be used to confirm the translocation and could be useful in cases where the *PML/RARA* transcript is poorly expressed.

FMS-Related Tyrosine Kinase 3 (*FLT3*), 13q12

Four classes of growth factor–receptor tyrosine kinases have been categorized (see Fig. 13.2). One class, represented by the ERBB (EGFR) family, was described in earlier sections. A second class includes dimeric receptors such as the insulin growth factor receptor (IGFR) and several proto-oncogenes. Members of the third class, including FMS, PDGFR, FLT1, and KIT, display five immunoglobulin-like domains in the extracellular region, and the catalytic domain is interrupted by a hydrophilic "interkinase" sequence of variable length. The fibroblast growth factor receptors (FGFR) represent the fourth class, which differ from the third class by having only three immunoglobulin-like domains in the extracellular region and a short kinase insert in the intracellular domain. FLT3 is a member of the third class of tyrosine-kinase receptors.

Particular mutations in the *FLT3* gene aberrantly activate the FLT3 kinase and predict prognosis in AML. These mutations include internal tandem duplications (ITDs) close to the transmembrane domain or point mutations affecting an aspartic acid residue in the kinase domain (D835 mutations). The ITD can easily be detected by PCR with primers flanking the potentially duplicated region. The size of the amplicon observed by agarose or capillary gel electrophoresis will increase in the event of an ITD.[130] D835 mutations can be detected by PCR-RFLP, where an *Eco*RV restriction site is destroyed by the presence of the mutation or by RT-PCR with FRET probes.[131] In performing these assays, it is important to have adequate representation of tumor cells in the specimen to avoid false-negative results from the presence of an excess of normal cells.

As data from mutation analysis grow, the interpretation of mutations has become more interrelated. For AML prognosis, it is recommended to interpret *FLT3* mutation results with karyotyping and tests for other gene mutations, including nucleophosmin *(NPM)* insertion mutations and point mutations in the CCAAT/enhancer-binding protein alpha *(CEBPA)* and the isocitrate dehydrogenase genes, *IDH1* and *IDH2*. If the karyotype is normal with an *NPM1* mutation but no *FLT3* ITD, or with a *CEBPA* mutation, the prognosis is favorable, similar to that of patients with chromosome

16 inversions or a t(8;21) translocation. With an *FLT3* ITD mutation and normal karyotype, the prognosis is less favorable.[132] *IDH* mutations are potential therapeutic targets.[133]

Janus Kinase 2 *(JAK2)*, 9p24

The *JAK2* gene codes for a kinase enzyme that phosphorylates several Signal Transducer and Activator of Transcription *(STAT)* gene products, bringing about cellular responses. A high proportion of patients with polycythemia vera, essential thrombocythemia, or idiopathic myelofibrosis carry a dominant mutation causing a valine-to-phenylalanine amino acid substitution at position 617 of the Jak2 protein (V617F). Detection of this mutation aids in the diagnosis of these myeloproliferative disorders. A convenient method of detection is to use multiplex sequence-specific PCR. In this assay, four primers are used, one forward and one reverse primer flanking the region of the mutation, one forward primer ending at the mutation site complementary to the normal base, and a fourth reverse primer also ending at the site of the mutation but complementary to the mutant base. The mutation is revealed as the product of the fourth reverse primer and the outer forward primer (see Chapter 8, Fig. 8.18). In some laboratories, white blood cells are fractionated in order to perform JAK2 testing specifically on the granulocyte fraction. The mutant/normal ratio is important, based on more recent data suggesting that the heterozygous/homozygous state of the V617F mutation has implications regarding which myeloproliferative disorder is present.[134] Furthermore very low levels of V617F may be found in normal individuals.

Four mutations affecting exon 12 of *JAK2* have been identified in some V617F-negative patients: F537-K539delinsL, K539L, H538QK539L, and N542E543del. These mutations are not located in the kinase domain of the Jak2 protein, but they do promote higher signaling in the cell than does the V617F mutation. Although the exon 12 mutations are not found in essential thrombocythemia, mutations in the myeloproliferative leukemia *(MPL)* gene may be present, suggesting a molecular differentiation between polycythemia and thrombocythemia. These mutations are detected using direct sequencing, although other methods, such as nonisotopic RNase cleavage assay (NIRCA), have been reported for *JAK2*.[135]

Mutation Spectra

Tumor-suppressor genes and oncogenes are disrupted by numerous genetic events. A number of specific abnormalities are associated with particular diseases (Table 13.7). These chromosome irregularities are targets for diagnosis of the associated diseases. The growing number of clinically significant variants has driven the use of high-throughput methods such as array analyses and NGS for more comprehensive information and more accurate interpretations.

Many molecular oncology tests have been developed by individual laboratories from published or original procedures rather than purchased as a set of reagents or in kit form from a commercial source. Often the final details of these procedures are determined empirically so that, in detail, a test procedure can differ from one laboratory to another. Even after the test procedure is established, troubleshooting performed as the procedure is put into use on a routine basis may further modify the procedure. Some reactions that work well for short-term research purposes may prove to be less consistent under the demands of the clinical laboratory setting. Even with high-throughput sequencing, the allelic fraction of mutations/normal sequence and patient characteristics (age, gender, type and location of tumor) have to be considered.

Biotechnology is rapidly developing standard reagent sets and sequencing panels (primers) for the most popular tests, but these may also differ from one supplier to another. Furthermore, due to market demands or new approaches such as "liquid biopsies," test reagent kits may be modified or discontinued. If replacement reagents are available, they may not be identical to those previously used. Ongoing tests then have to be optimized. This can be a concern where turnaround times are critical.

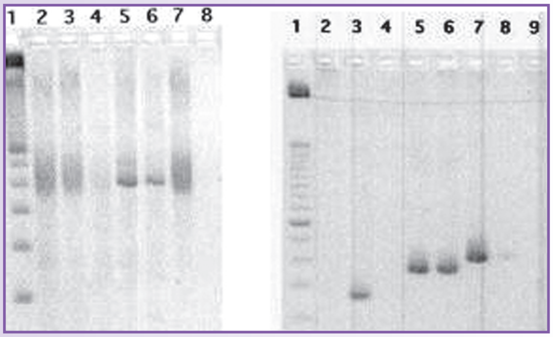
It then becomes the responsibility of the technologist to perform and monitor tests on a regular basis to maintain consistency and accuracy of results. The technologist who understands the biochemistry and molecular biology of these tests will be better able to respond to these problems. In addition, with the evolution of the sciences, a knowledgeable technologist can better recognize significant discoveries that offer the potential for test improvement.

TABLE 13.7 Chromosomal Abnormalities Associated With Leukemias and Lymphomas

Disease	Chromosomal Mutation
Pre-B acute lymphoblastic leukemia	t(1;19)
Acute lymphocytic leukemia	t(4;11), t(11;14), t(9;22), del(12)(p11) or (p11p13) or t(12)(p11), i(17q), t(9;22), t(12;21), t(8;14), t(2;8), t(8;22), t(11q)
B-cell leukemia	t(2;8), t(8;14), t(8;22), t(11;14)
Acute T-lymphocytic leukemia	t(11;14), del(9)(p21 or 22)
Acute myeloid leukemia/myelodysplastic syndrome	t(11q23)-multiple partners
Acute myeloid leukemia (M2)	t(8;21), t(6;9)
Acute promyelocytic leukemia (M3)	t(15;17), 14q+
Acute myelomonocytic leukemia (M4)	t(11;21), inv(16)(p13q22)
Acute monocytic leukemia (M5)	t(9;11), del(11)(q23), t(11q23)-multiple partners
Chronic myelogenous leukemia	t(9;22), t(11;22), +8, +12, i(17q)
Acute nonlymphocytic leukemia	t(8;21), -Y
Chronic lymphocytic leukemia	14q+, +12, t(14;19), del(11)(q22), del(13q), del(17)(p13)
T-chronic lymphocytic leukemia	Inv(14)(q11q32) or t(14;14)(q11;q32)
Burkitt lymphoma	t(8;14), t(2;8), t(8;22)
Diffuse large B-cell lymphoma	t(3q27), t(14;18), t(8;14)
T-cell lymphoma	t(8;14)
Follicular lymphoma	t(14;18), t(8;14)
Mantle cell lymphoma	t(11;14)
Multiple myeloma	t(14q32), 14q+
Myeloproliferative/myelodysplastic disease	del(5)(q12q33), -7 or del(7)(q22), +8, +9
Myeloproliferative/myeloblastic disease	del(13)(q12 or q14), +21
Hairy cell leukemia	14q+
Waldenström macroglobulinemia	14q+
Mucosa-associated lymphoid tissue lymphoma	t(11;18), t(14;18), t(1;14)
Polycythemia vera	del(20)(q11.2q13.3)

Case Study 13.1

A 40-year-old woman with a history of non-Hodgkin's lymphoma reported to her physician for follow-up testing. Her complete blood count (CBC) was normal, including a white blood cell (WBC) count of 11,000/µL. Morphological studies on a bone marrow biopsy and a bone marrow aspirate revealed several small aggregates of mature and immature lymphocytes. Flow-cytometry studies were difficult to interpret because there were too few B cells in the bone marrow aspirate specimen. No chromosomal abnormalities were detected by cytogenetics. A bone marrow aspirate tube was also sent for molecular analysis—namely, immunoglobulin heavy-chain gene rearrangement and t(14;18) gene translocation analysis by PCR. The immunoglobulin heavy-chain gene rearrangement results are shown in lane 4 of the following gel image:



Immunoglobulin heavy-chain gene rearrangement results *(left)*: lane 1, molecular-weight marker; lanes 2 to 4, patient specimens; lane 5, positive control; lane 6, sensitivity; lane 7, negative control; lane 8, reagent blank. t(14;18) gene translocation test *(right)*: lane 1, molecular-weight marker; lanes 2 to 6, patient specimens; lane 7, positive control; lane 8, sensitivity; lane 9, negative control.

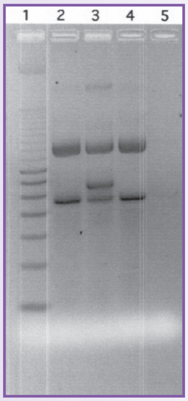
QUESTIONS:

1. *Does this patient have an amplifiable gene rearrangement?*
2. *Are the translocation results at right consistent with those of cytogenetics?*
3. *What control is missing from the translocation analysis?*

Case Study 13.2

A 54-year-old woman with thrombosis, a high platelet count (900,000/µL), and a decreased erythrocyte sedimentation rate was tested for polycythemia vera. Megakaryocyte clusters and pyknotic nuclear clusters were observed in a bone marrow biopsy. Overall cellularity was decreased.

In the CBC, white blood cell and neutrophil counts were normal. Iron stores were also in the normal range. A blood sample was submitted to the molecular pathology laboratory for JAK2 V617F mutation analysis. The results were reported as follows:



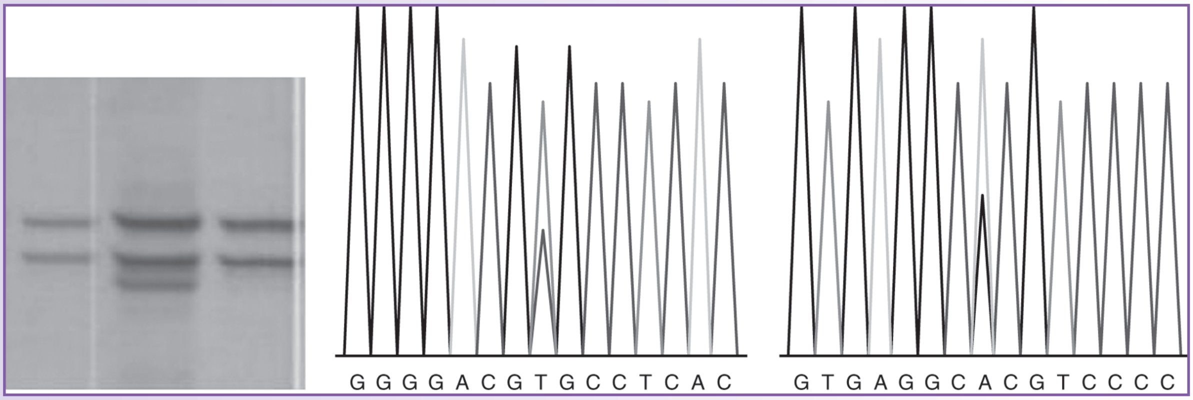
Agarose gel electrophoresis of SSP-PCR amplicons. Extension of a JAK2 V617F–specific primer produces a 270-bp band in addition to the normal 500- and 230-bp bands. Lane 1, molecular-weight markers; lane 2, patient specimen; lane 3, V617F mutant control; lane 4, normal control; lane 5, reagent blank.

QUESTIONS:

1. *What is the source of the two bands in the normal control in lane 4?*
2. *What is the source of the additional band in the positive control in lane 3?*
3. *Does this patient have the JAK2 V617F mutation?*

Case Study 13.3

Paraffin-embedded sections were submitted to the molecular diagnostics laboratory for p53 mutation analysis. The specimen was a small tumor (intraductal carcinoma in situ) discovered in a 55-year-old woman. Lymph nodes were negative. Slides stained with hematoxylin and eosin were examined for confirmation of the location of tumor cells on the sections. These cells were dissected from the slide. DNA isolated from the microdissected tumor cells was screened by SSCP for mutations in exons 4 to 9 of the p53 gene. The results for exon 5 were reported as follows:



Polyacrylamide gel electrophoresis with silver-stain detection of p53 exon 5 (*left*). Lane 1, normal; lane 2, patient; lane 3, normal. Direct sequencing (*right*) revealed a C→T (G→A) base change, resulting in an R→H amino acid change at position 175.

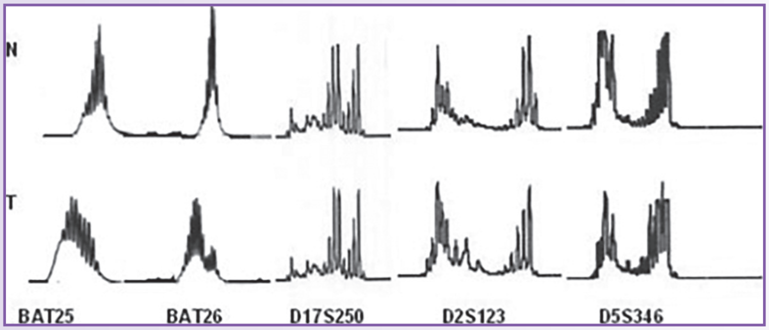
QUESTIONS:

1. *What information is gained from the SSCP results?*
2. *What is the source of the additional band in the patient SSCP lane?*
3. *What further confirmation/information is gained from direct sequencing?*

Case Study 13.4

Colon carcinoma and three polyps were resected in a right hemicolectomy of a 33-year-old man. Without a family history, the man's age and the location of the tumor warranted testing for HNPCC. Histological staining was negative for MSH2 protein. Paraffin sections and a blood sample were submitted to the molecular pathology laboratory for microsatellite instability testing. DNA was isolated from tumor cells dissected from four paraffin sections and from the patient's white blood cells. Both were amplified at the five microsatellite loci recommended by the National Cancer Institute. The results were reported as follows:

Case Study 13.4 *(Continued)*



MSI analysis of normal cells (N, *top row*) and tumor cells (T, *bottom row*) at the five NCI loci.

QUESTIONS:

1. *What are the indications for MSI analysis in this case?*
2. *What accounts for the difference in the peak patterns in the BAT25 and BAT26 loci?*
3. *How would you interpret the results of this analysis according to NCI recommendations?*

Case Study 13.5

A 23-year-old college student was sent to the student health office with a painful bruise that persisted for several weeks. He was referred to the local hospital, where the physician ordered a bone marrow biopsy and a CBC. His WBC count was 28,000/µL; red blood cells, 2 million/µL; hemoglobin 8, hematocrit 20, platelets, 85,000/µL. Neutrophils were 7%, lymphocytes were 5%, and blasts were 95%. The pathologist who examined the bone marrow biopsy requested flow-cytometry analysis of the aspirate. The cells expressed 84% CD20, 82% CD34, 92% HLA-DR, and 80% CD10/CD19. The results of these tests indicated a diagnosis of acute lymphoblastic leukemia. A blood specimen was sent to the cytogenetics laboratory for chromosomal analysis. Twenty metaphases examined had a normal 46,XY chromosomal complement. Interphase FISH was negative for t(9;22) in 500 nuclei.

QUESTIONS:

1. *What is the significance of the cytogenetic test results?*
2. *What other molecular abnormalities might be present in this type of tumor?*

STUDY QUESTIONS

1. What are the two important checkpoints in the cell division cycle that are crossed when the regulation of the cell division cycle is affected?

2. An EWS-FLI-1 mutation was detected in a solid tumor by RT-PCR. Which of the following does this result support?

   a. Normal tissue
   b. Ewing sarcoma
   c. Inherited breast cancer
   d. Microsatellite instability

3. Mutation detection, even by sequencing, is not definitive with a negative result. Why?

4. A PCR test for the BCL-2 translocation is performed on a patient with suspected follicular lymphoma. The results show a bright band at about 300 bp for this patient. How would you interpret these results?

5. Which of the following misinterpretations would result from PCR contamination?
   - a. False positive for the t(15;17) translocation
   - b. False negative for the t(15;17) translocation
   - c. False negative for a gene rearrangement

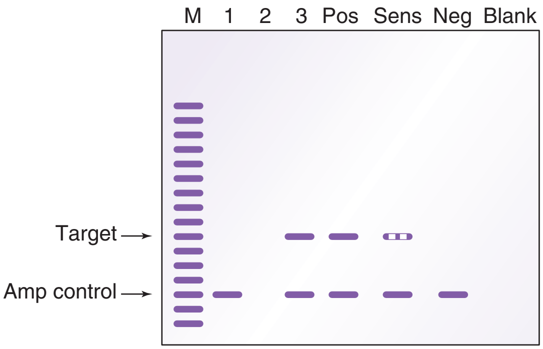
6. After amplification of the t(12;21) breakpoint by qRT-PCR, what might be the explanation for each of the following observations? (Assume that positive and amplification controls and a reagent blank control are included in the run.)
   - a. The gel is blank (no bands, no molecular-weight standard).
   - b. Only the molecular-weight standard is visible.
   - c. The molecular-weight standard is visible; there are bands in every lane at 200 bp, even in the reagent blank lane.

7. What is observed on a Southern blot for gene rearrangement in the case of a positive result?
   - a. No bands
   - b. Germline bands plus rearranged bands
   - c. Smears
   - d. Germline bands only

8. Cyclin D1 promotes passage of cells through the G1-to-S checkpoint. What test detects translocation of this gene to chromosome 14?
   - a. t(14;18) translocation analysis *(BCL2, IGH)*
   - b. t(15;17) translocation analysis *(PML/RARA)*
   - c. t(11;14) translocation analysis *(BCL1/IGH)*
   - d. t(8;14) translocation analysis *(MYC/IGH)*

9. Why is the Southern blot procedure superior to the PCR procedure for detecting clonality in some cases?
   - a. Southern blot requires less sample DNA than does PCR.
   - b. The PCR procedure cannot detect certain gene rearrangements that are detectable by Southern blot.
   - c. Southern blot results are easier to interpret than PCR results.
   - d. PCR results are not accepted by the College of American Pathologists.

10. Interpret the following results from a translocation assay.



Are the samples positive, negative, or indeterminate?
Sample 1:
Sample 2:
Sample 3:

11. Which of the following predicts the efficacy of EGFR tyrosine-kinase inhibitors?
    - a. Overexpression of EGFR protein
    - b. EGFR-activating mutations
    - c. Patient gender
    - d. Stage of disease

12. What is the advantage of macrodissection in testing for tumor-specific molecular markers from paraffin-embedded formalin-fixed tissue sections?

13. Why are KRAS- and BRAF-activating mutations almost always exclusive of one another?

14. What enzyme is responsible for continued sequence changes in the immunoglobulin heavy-chain gene variable region after gene rearrangement has occurred?

15. Why are translocation-based PCR tests more sensitive than IgH, IgL, or T-cell receptor gene-rearrangement tests?

References

1. Institute NC. National Cancer Institute sponsored study of classifications of non-Hodgkin's lymphomas: summary and description of a working formulation for clinical usage. The Non-Hodgkin's Lymphoma Pathologic Classification Project. *Cancer* 1982;49:2112–2135.
2. Pileri S, Agostinelli C, Sabattini E, Bacci F, Sagramoso C, Pileri A Jr., Falini B, Piccaluga PP. Lymphoma classification: the quiet after the storm. *Seminars in Diagnostic Pathology* 2011;28:113–123.
3. Rosolen A, Perkins SL, Pinkerton CR, Guillerman RP, Sandlund JT, Patte C, Reiter A, Cairo MS. Revised international pediatric non-Hodgkin lymphoma staging system. *Journal of Clinical Oncology* 2015;33:2112–2118.
4. Press M, Slamon DJ, Flom KJ, Park J, Zhou JY, Bernstein L. Evaluation of *HER-2/neu* gene amplification and overexpression: comparison of frequently used assay methods in a molecularly characterized cohort of breast cancer specimens. *Journal of Clinical Oncology* 2002;20:3095–3105.
5. Wirtz R, Sihto H, Isola J, Heikkilä P, Kellokumpu-Lehtinen PL, Auvinen P, Turpeenniemi-Hujanen T, Jyrkkiö S, Lakis S, Schlombs K, Laible M, Weber S, Eidt S, Sahin U, Joensuu H. Biological subtyping of early breast cancer: a study comparing RT-qPCR with immunohistochemistry. *Breast Cancer Research and Treatment* 2016;157:437–446.
6. Sharifi N, Salmaninejad A, Ferdosi S, Bajestani AN, Khaleghiyan M, Estiar MA, Jamali M, Nowroozi MR, Shakoori A. *HER2* gene amplification in patients with prostate cancer: evaluating a CISH-based method. *Oncology Letters* 2016;12:4651–4658.
7. Arnould L, Roger P, Macgrogan G, Chenard MP, Balaton A, Beauclair S, Penault-Llorca F. Accuracy of *HER2* status determination on breast core-needle biopsies (immunohistochemistry, FISH, CISH and SISH vs FISH). *Modern Pathology* 2010;25:675–682.
8. Chae YK, Arya A, Chiec L, Shah H, Rosenberg A, Patel S, Raparia K, Choi J, Wainwright DA, Villaflor V, Cristofanilli M, Giles F. Challenges and future of biomarker tests in the era of precision oncology: can we rely on immunohistochemistry (IHC) or fluorescence in situ hybridization (FISH) to select the optimal patients for matched therapy? *Oncogene* 2017;8:100863–100898.
9. Younes M. Is immunohistochemistry for epidermal growth factor receptor expression a poor predictor of response to epidermal growth factor receptor-targeted therapy? *Journal of Clinical Oncology* 2005;23:923.
10. Murthy S, Magliocco AM, Demetrick DJ. Copy number analysis of *c-erb-B2 (HER-2/neu)* and topoisomerase IIa genes in breast carcinoma by quantitative real-time polymerase chain reaction using hybridization probes and fluorescence in situ hybridization. *Archives of Pathology & Laboratory Medicine* 2005;129:39–46.
11. Wang J, Du Q, Li C. Bioinformatics analysis of gene expression profiles to identify causal genes in luminal B2 breast cancer. *Oncology Letters* 2017;14:7880–7888.
12. Lynch T, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, Harris PL, Haserlat SM, Supko JG, Haluska FG, Louis DN, Christiani DC, Settleman J, Haber DA. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *New England Journal of Medicine* 2004;350:2129–2139.
13. Paez J, Janne PA, Lee JC, Tracy S, Greulich H, Gabriel S, Herman P, Kaye FJ, Lindeman N, Boggon TJ, Naoki K, Sasaki H, Fujii Y, Eck MJ, Sellers WR, Johnson BE, Meyerson M. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* 2004;304:1497–1500.
14. Tsao M-S, Sakurada A, Cutz J-C, Zhu CQ, Kamel-Reid S, Squire J, Lorimer I, Zhang T, Liu N, Daneshmand M, Marrano P, da Cunha Santos G, Lagarde A, Richardson F, Seymour L, Whitehead M, Ding K, Pater J, Shepherd FA. Erlotinib in lung cancer—molecular and clinical predictors of outcome. *New England Journal of Medicine* 2005;353:133–144.
15. Shepherd F, Pereira JR, Ciuleanu T, Tan EH, Hirsh V, Thongprasert S, Campos D, Maoleekoonpiroj S, Smylie M, Martins R, van Kooten M, Dediu M, Findlay B, Tu D, Johnston D, Bezjak A, Clark G, Santabárbara P, Seymour L. Erlotinib in previously treated non–small-cell lung cancer. *New England Journal of Medicine* 2005;353:123–132.
16. Marchetti A, Martella C, Felicioni L, Barassi F, Salvatore S, Chella A, Camplese PP, Iarussi T, Mucilli F, Mezzetti A, Cuccurullo F, Sacco R, Buttitta F. EGFR mutations in non-small-cell lung cancer: analysis of a large series of cases and development of a rapid and sensitive method for diagnostic screening with potential implications on pharmacologic treatment. *Journal of Clinical Oncology* 2005;23:857–865.
17. Pan Q, Pao W, Ladanyi M. Rapid polymerase chain reaction-based detection of epidermal growth factor receptor gene mutations in lung adenocarcinomas. *Journal of Molecular Diagnostics* 2005;7:396–403.
18. Illei P, Belchis D, Tseng LH, Nguyen D, De Marchi F, Haley L, Riel S, Beierl K, Zheng G, Brahmer JR, Askin FB, Gocke CD, Eshleman JR, Forde PM, Lin MT. Clinical mutational profiling of 1006 lung cancers by next generation sequencing. *Oncotarget* 2017;20:96684–96696.
19. Mehrad M, Roy S, Bittar HT, Dacic S. Next-generation sequencing approach to non-small cell lung carcinoma yields more actionable alterations. *Archives of Pathology & Laboratory Medicine* 2017;142(3):353–357.
20. Taron M, Ichinose Y, Rosell R, Mok T, Massuti B, Zamora L, Mate JL, Manegold C, Ono M, Queralt C, Jahan T, Sanchez JJ, Sanchez-Ronco M, Hsue V, Jablons D, Sanchez JM, Moran T. Activating mutations in the tyrosine kinase domain of the epidermal growth factor receptor are associated with improved survival in gefitinib-treated chemorefractory lung adenocarcinomas. *Clinical Cancer Research* 2005;11:5878–5885.
21. Han S, Hwang PG, Chun DH, Kim DW, Im SA, Kim YT, Kim TY, Heo DS, Bang YJ, Kim NK. Epidermal growth factor receptor (EGFR) downstream molecules as response predictive markers for gefitinib (Iressa, ZD1839) in chemotherapy-resistant non-small cell lung cancer. *International Journal of Cancer* 2005;113:109–115.
22. Jassem J, Jassem E, Jakobkiewicz-Banecka J, Rzyman W, Badzio A, Dziadziuszko R, Kobierska-Gulida G, Szymanowska A, Skrzypski M, Zylicz M. *P53* and *K-ras* mutations are frequent

events in microscopically negative surgical margins from patients with nonsmall cell lung carcinoma. *Cancer* 2004;100(9): 1951–1960.

23. Tonon G, Wong K-K, Maulik G, Brennan C, Feng B, Zhang Y, Khatry DB, Protopopov A, You MJ, Aguirre AJ, Martin ES, Yang Z, Ji H, Chin L, DePinho RA. High-resolution genomic profiles of human lung cancer. *Proceedings of the National Academy of Sciences* 2005;102:9625–9630.

24. Tramm T, Kyndi M, Sørensen FB, Overgaard J, Alsner J. Influence of intra-tumoral heterogeneity on the evaluation of *BCL2,* E-cadherin, EGFR, EMMPRIN, and Ki-67 expression in tissue microarrays from breast cancer. *Acta Oncology* 2017;57(1): 102–106.

25. Chen I, Raymond VM, Geis JA, Collisson EA, Jensen BV, Hermann KL, Erlander MG, Tempero M, Johansen JS. Ultrasensitive plasma ctDNA KRAS assay for detection, prognosis, and assessment of therapeutic response in patients with unresectable pancreatic ductal adenocarcinoma. *Oncotarget* 2017;8: 97769–97786.

26. Illei P, Belchis D, Tseng LH, Nguyen D, De Marchi F, Haley L, Riel S, Beierl K, Zheng G, Brahmer JR, Askin FB, Gocke CD, Eshleman JR, Forde PM, Lin MT. Clinical mutational profiling of 1006 lung cancers by next generation sequencing. *Oncotarget* 2017;8:96684–96696.

27. Reiman A, Kikuchi H, Scocchia D, Smith P, Tsang YW, Snead D, Cree IA. Validation of an NGS mutation detection panel for melanoma. *BMC Cancer* 2017;17:150–156.

28. Li W QT, Guo L, Ying J. Major challenges related to tumor biological characteristics in accurate mutation detection of colorectal cancer by next-generation sequencing. *Cancer Letters* 2017;410:92–99.

29. Machado I, Navarro L, Pellin A, Navarro S, Agaimy A, Tardío JC, Karseladze A, Petrov S, Scotlandi K, Picci P, Llombart-Bosch A. Defining Ewing and Ewing-like small round cell tumors (SRCT): the need for molecular techniques in their categorization and differential diagnosis. *Annals of Diagnostic Pathology* 2016;22:25–32.

30. Noguera R. Cytogenetics and tissue culture of small round cell tumors of bone and soft tissue. *Seminars in Diagnostic Pathology* 1996;13:171–183.

31. Akerman M, Dreinhöfer K, Rydholm A, Willén H, Mertens F, Mitelman F, Mandahl N. Cytogenetic studies on fine-needle aspiration samples from osteosarcoma and Ewing's sarcoma. *Diagnostic Cytopathology* 1996;15:17–22.

32. Qian X, Jin L, Shearer B, Ketterling RP, Jalal SM, Lloyd RV. Molecular diagnosis of Ewing's sarcoma/primitive neuroectodermal tumor in formalin-fixed paraffin-embedded tissues by RT-PCR and fluorescence in situ hybridization. *Diagnostic Molecular Pathology* 2005;14:23–28.

33. Warren M, Weindel M, Ringrose J, Venable C, Reyes A, Terashima K, Rao P, Chintagumpala M, Hicks MJ, Lopez-Terrada D, Lu XY. Integrated multimodal genetic testing of Ewing sarcoma–a single-institution experience. *Human Pathology* 2013;44:2010–2019.

34. Mohamed M, Gonzalez D, Fritchie KJ, Swansbury J, Wren D, Benson C, Jones RL, Fisher C, Thway K. Desmoplastic small round cell tumor: evaluation of reverse transcription-polymerase chain reaction and fluorescence in situ hybridization as ancillary molecular diagnostic techniques. *Virchows Archiv* 2017;471:631–640.

35. Hayashi M, Chu D, Meyer CF, Llosa NJ, McCarty G, Morris CD, Levin AS, Wolinsky JP, Albert CM, Steppan DA, Park BH, Loeb DM. Highly personalized detection of minimal Ewing sarcoma disease burden from plasma tumor DNA. *Cancer* 2016;122:3015–3023.

36. Panagopoulos I, Mertens F, Isaksson M, Limon J, Gustafson P, Skytting B, Akerman M, Sciot R, Dal Cin P, Samson I, Iliszko M, Ryoe J, Dubiec-Rychter M, Szadowska A, Brosja O, Larsson O, Rydholm A, Mandahl N. Clinical impact of molecular and cytogenetic findings in synovial sarcoma. *Genes Chromosomes Cancer* 2001;31:362–372.

37. Owosho A, Estilo CL, Rosen EB, Yom SK, Huryn JM, Antonescu CR. A clinicopathologic study on SS18 fusion positive head and neck synovial sarcomas. *Oral Oncology* 2017;66:46–51.

38. Barr F, Galili N, Holick J, Biegel JA, Rovera G, Emanuel BS. Rearrangement of the *PAX3* paired box gene in the paediatric solid tumour alveolar rhabdomyosarcoma. *Nature Genetics* 1993;3:113–117.

39. Tassabehji M, Newton VE, Liu X-Z, Brady A, Donnai D, Krajewska-Walasek M, Murday V, Norman A, Obersztyn E, Reardon W, Rice JC, Trembath R, Wieacker P, Whiteford M, Winter R, Read AP. The mutational spectrum in Waardenburg syndrome. *Human Molecular Genetics* 1995;4:2131–2137.

40. Niu Z, Li J, Tang F, Sun J, Wang X, Jiang L, Mei L, Chen H, Liu Y, Cai X, Feng Y, He C. Identification and functional analysis of a novel mutation in the *PAX3* gene associated with Waardenburg syndrome type I. *Gene* 2018;642:362–366.

41. Jin L, Majerus J, Oliveira A, Inwards CY, Nascimento AG, Burgart LJ, Lloyd RV. Detection of fusion gene transcripts in fresh-frozen and formalin-fixed paraffin-embedded tissue sections of soft-tissue sarcomas after laser capture microdissection and RT-PCR. *Diagnostic Molecular Pathology* 2003;12:224–230.

42. Tiffin N, Williams RD, Shipley J, Pritchard-Jones K. PAX7 expression in embryonal rhabdomyosarcoma suggests an origin in muscle satellite cells. *British Journal of Cancer* 2003;89: 327–332.

43. Xie Z, Babiceanu M, Kumar S, Jia Y, Qin F, Barr FG, Li H. Fusion transcriptome profiling provides insights into alveolar rhabdomyosarcoma. *Proceedings of the National Academy of Sciences* 2016;113:13126–13131.

44. Bykov V, Eriksson SE, Bianchi J, Wiman KG. Targeting mutant p53 for efficient cancer therapy. *Nature Review Cancer* 2017;18(2):89–102.

45. Kyzas P, Loizou KT, Ioannidis JP. Selective reporting biases in cancer prognostic factor studies. *Journal of the National Cancer Institute* 2005;97:1043–1055.

46. Lohmann D, Ruhri C, Schmitt M, Graeff H, Hofler H. Accumulation of p53 protein as an indicator for p53 gene mutation in breast cancer. Occurrence of false-positives and false-negatives. *Diagnostic Molecular Pathology* 1993;2:36–41.

47. Visscher D, Sarkar FH, Shimoyama RK, Crissman JD. Correlation between p53 immunostaining patterns and gene sequence mutations in breast carcinoma. *Diagnostic Molecular Pathology* 1996;5:187–193.

48. Girard P-M, Foray N, Stumm M, Waugh A, Riballo E, Maser RS, Phillips WP, Petrini J, Arlett CF, Jeggo PA. Radiosensitivity in Nijmegen Breakage Syndrome cells is attributable to a repair defect and not cell cycle checkpoint defects. *Cancer Research* 2000;60(17):4881–4888.
49. Gomez-Cabello D, Jimeno S, Fernandez-Avila MJ, Huertas P. New tools to study DNA double-strand break repair pathway choice. *PLoS One* 2013;8:e77206.
50. Gavish-Izakson M, Velpula BB, Elkon R, Prados-Carvajal R, Barnabas GD, Ugalde AP, Agami R, Geiger T, Huertas P, Ziv Y, Shiloh Y. Nuclear poly(A)-binding protein 1 is an ATM target and essential for DNA double-strand break repair. *Nucleic Acids Research* 2017;46(2):730–747.
51. Castilla L, Couch FJ, Erdos MR, Hoskins KF, Calzone K, Garber JE, Boyd J, Lubin MB, Deshano ML, Brody LC. Mutations in the *BRCA1* gene in families with early-onset breast and ovarian cancer. *Nature Genetics* 1994;8:387–391.
52. Smith S, DiCioccio RA, Struewing JP, Easton DF, Gallion HH, Albertsen H, Mazoyer S, Johansson B, Steichen-Gersdorf E, Stratton M. Localisation of the breast-ovarian cancer susceptibility gene *(BRCA1)* on 17q12-21 to an interval of < or = 1 cM. *Genes Chromosomes Cancer* 1994;10:71–76.
53. Stratton M, Ford D, Neuhasen S, Seal S, Wooster R, Friedman LS, King MC, Egilsson V, Devilee P, McManus R. Familial male breast cancer is not linked to the *BRCA1* locus on chromosome 17q. *Nature Genetics* 1994;7:103–107.
54. Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, Collins N, Gregory S, Gumbs C, Micklem G. Identification of the breast cancer susceptibility gene *BRCA2*. *Nature* 1995;378:789–792.
55. Ohmoto A, Yachida S. Current status of poly(ADP-ribose) polymerase inhibitors and future directions. *Oncotargets and Therapy* 2017;10:5195–5208.
56. Ferrara R, Simionato F, Ciccarese C, Grego E, Cingarlini S, Iacovelli R, Bria E, Tortora G, Melisi D. The development of PARP as a successful target for cancer therapy. *Expert Reviews in Anticancer Therapy* 2017;18:161–175.
57. Tarade D, Ohh M. The HIF and other quandaries in VHL disease. *Oncogene* 2017;37(2):139–147.
58. Trop-Steinberg S, Azar Y. Is *myc* an important biomarker? *Myc* expression in immune disorders and cancer. *American Journal of Medical Science* 2018;355:67–75.
59. Harada K, Ferdous T, Cui D, Kuramitsu Y, Matsumoto T, Ikeda E, Okano H, Ueyama Y. Induction of artificial cancer stem cells from tongue cancer cells by defined reprogramming factors. *BMC Cancer* 2016;16:548.
60. Cai Q, Medeiros LJ, Xu X, Young KH. *Myc*-driven aggressive B-cell lymphomas: biology, entity, differential diagnosis and clinical management. *Oncotarget* 2015;6:38591–38616.
61. Layfield L, Willmore-Payne C, Shimada H, Holden JA. Assessment of NMYC amplification: a comparison of FISH, quantitative PCR monoplexing and traditional blotting methods used with formalin-fixed, paraffin-embedded neuroblastomas. *Analytical and Quantitative Cytology and Histology* 2005;27: 5–14.
62. Bolk G, Salomon R, Pelet A, Angrist M, Amiel J, Fornage M, Attie-Bitach T, Olson JM, Hofstra R, Buys C, Steffann J, Munnich A, Lyonnet S, Chakravarti A. Segregation at three loci explains familial and population risk in Hirschsprung disease. *Nature Genetics* 2002;31:89–93.
63. Zhang Y, Zheng D, Li Y, Pan Y, Sun Y, Chen H. Comprehensive investigation of clinicopathologic features, oncogenic driver mutations and immunohistochemical markers in peripheral lung squamous cell carcinoma. *Journal of Thoracic Diseases* 2017;9(11):4434–4440.
64. Griffin R, Ramirez RA. Molecular targets in non-small cell lung cancer. *The Ochsner Journal* 2017;17:388–392.
65. Kanth P, Grimmett J, Champine M, Burt R, Samadder NJ. Hereditary colorectal polyposis and cancer syndromes: a primer on diagnosis and management. *American Journal of Gastroenterology* 2017;112:1509–1525.
66. Blount J, Prakash A. The changing landscape of Lynch syndrome due to *PMS2* mutations. *Clinical Genetics* 2017;94(1):61–69.
67. Boland C, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, Meltzer SJ, Rodriguez-Bigas MA, Fodde R, Ranzani GN, Srivastava S. A National Cancer Institute Workshop on microsatellite instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Research* 1998;58:5248–5257.
68. Libera L, Sahnane N, Carnevali IW, Cimetti L, Cerutti R, Chiaravalli AM, Riva C, Tibiletti MG, Sessa F, Furlan D. Microsatellite analysis of sporadic and hereditary gynaecological cancer in routine diagnostics. *Journal of Clinical Pathology* 2017;70:792–797.
69. Coelho H, Jones-Hughes T, Snowsill T, Briscoe S, Huxley N, Frayling IM, Hyde C. A systematic review of test accuracy studies evaluating molecular micro-satellite instability testing for the detection of individuals with Lynch syndrome. *BMC Cancer* 2017;17:836.
70. Cohen R, Buhard O, Cervera P, Hain E, Dumont S, Bardier A, Bachet JB, Gornet JM, Lopez-Trabada D, Dumont S, Kaci R, Bertheau P, Renaud F, Bibeau F, Parc Y, Vernerey D, Duval A, Svrcek M, André T. Clinical and molecular characterisation of hereditary and sporadic metastatic colorectal cancers harbouring microsatellite instability/DNA mismatch repair deficiency. *European Journal of Cancer* 2017;86:266–274.
71. Tutlewska K, Lubinski J, Kurzawski G. Germline deletions in the *EPCAM* gene as a cause of Lynch syndrome—literature review. *Hereditary Cancer in Clinical Practice* 2013;11:1–9.
72. Usha L, Dewdney SB, Buckingham LE. Tumor screening and DNA testing in the diagnosis of Lynch syndrome. *Journal of the American Medical Association* 2016;316:93–94.
73. Rubocki R, Duffy KJ, Shepard KL, McCue BJ, Shepherd SJ, Wisecarver JL. Loss of heterozygosity detected in a short tandem repeat (STR) locus commonly used for human DNA identification. *Journal of Forensic Science* 2000;45:1087–1089.
74. Perakis S, Speicher MR. Emerging concepts in liquid biopsies. *BMC Medicine* 2017;15:1–12.
75. Raja T, Warrier NK. Epidermal growth factor receptor mutation testing: from conventional to real-time diagnosis of lung cancer. *Indian Journal of Cancer* 2017;54:s9–s14.
76. Sands J, Li Q, Hornberger J. Urine circulating-tumor DNA (ctDNA) detection of acquired EGFR T790M mutation in

non-small-cell lung cancer: an outcomes and total cost-of-care analysis. *Lung Cancer* 2017;110:19–25.

77. Franovic A, Raymond VM, Erlander MG, Reckamp KL. Urine test for EGFR analysis in patients with non-small cell lung cancer. *Journal of Thoracic Diseases* 2017;9:s1323–s1331.

78. Cho E, Wendel M, Luttgen M, Yoshioka C, Marrinucci D, Lazar D, Schram E, Nieva J, Bazhenova L, Morgan A, Ko AH, Korn WM, Kolatkar A, Bethel K, Kuhn P. Cho EH, Wendel M, Luttgen M, Yoshioka C, Marrinucci D, Lazar D, et al. Characterization of circulating tumor cell aggregates identified in patients with epithelial tumors. *Physical Biology* 2012;9:1–13.

79. Weichhold G, Ohnheiser R, Zachau HG. The human immunoglobulin kappa locus consists of two copies that are organized in opposite polarity. *Genomics* 1993;16:503–511.

80. Barbie V, Lefranc M-P. The human immunoglobulin kappa variable (IGKV) genes and joining (IGKJ) segments. *Experimental and Clinical Immunogenetics* 1998;15:171–183.

81. Frippiat J-P, Williams SC, Tomlinson IM, Cook GP, Cherif D, Le Paslier D, Collins JE, Dunham I, Winter G, Lefranc M-P. Organization of the human immunoglobulin lambda light-chain locus on chromosome 22q11.2. *Human Molecular Genetics* 1995;4:983–991.

82. Meier J, Lewis SM. P nucleotides in V(D)J recombination: a fine-structure analysis. *Molecular and Cellular Biology* 1993;13:1078–1092.

83. Yuen E, Brown RD. Southern blotting of IgH rearrangements in B-cell disorders. *Methods in Molecular Medicine* 2005;113:85–103.

84. van Dongen J, Langerak AW, Bruggemann M, Evans PA, Hummel M, Lavender FL, Delabesse E, Davi F, Schuuring E, Garcia-Sanz R, van Krieken JH, Droese J, Gonzalez D, Bastard C, White HE, Spaargaren M, Gonzalez M, Parreira A, Smith JL, Morgan GJ, Kneba M, Macintyre EA. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. *Leukemia* 2003;12:2257–2317.

85. Lu C, He Q, Zhu W, Fu C, Zhou J, Tao Y, Liu S, Xiao D. The value of detecting immunoglobulin gene rearrangements in the diagnosis of B-cell lymphoma. *Oncotarget* 2017;8:77009–77019.

86. Jonsson S, Sveinbjornsson G, de Lapuente Portilla AL, Swaminathan B, Plomp R, Dekkers G, Ajore R, Ali M, Bentlage AEH, Elmér E, Eyjolfsson GI, Gudjonsson SA, Gullberg U, Gylfason A, Halldorsson BV, Hansson M, Holm H, Johansson Å, Johnsson E, Jonasdottir A, Ludviksson BR, Oddsson A, Olafsson I, Olafsson S, Sigurdardottir O, Sigurdsson A, Stefansdottir L, Masson G, Sulem P, Wuhrer M, Wihlborg AK, Thorleifsson G, Gudbjartsson DF, Thorsteinsdottir U, Vidarsson G, Jonsdottir I, Nilsson B, Stefansson K. Identification of sequence variants influencing immunoglobulin levels. *Nature Genetics* 2017;49:1182–1191.

87. Smol T, Daudignon A. Comparison of IGH profile signals using t(4;14) and IGH break-apart probes by FISH in multiple myeloma. *Cytogenetic and Genome Research* 2017;153(1):18–21.

88. Ono M, Nose M. Persistent expression of an unproductive immunoglobulin heavy chain allele with DH-JH-gamma configuration in peripheral tissues. *APMIS* 2007;115:1350–1356.

89. Gonzalez D, Balanzategui A, Garcia-Sanz R, Guiterrez N, Seabra C, van Dongen JIM, Gozalez M, San Miguel JF. Incomplete DJH gene rearrangements of the *IgH* gene are frequent in multiple myeloma patients: immunobiological characteristics and clinical applications. *Leukemia* 2003;17:1398–1403.

90. Gonzalez D, Garcia-Sanz R. Incomplete DJH rearrangements. *Methods in Molecular Medicine* 2005;113:165–173.

91. Beishuizen A, Verhoeven MA, Mol EJ, van Dongen JJ. Detection of immunoglobulin kappa light-chain gene rearrangement patterns by Southern blot analysis. *Leukemia* 1994;8:2228–2236.

92. van der Velden V, Willemse MJ, van der Schoot CE, Hohlen K, van Wering ER, van Dongen JJ. Immunoglobulin kappa deleting element rearrangements in precursor-B acute lymphoblastic leukemia are stable targets for detection of minimal residual disease by real-time PCR. *Leukemia* 2002;16:928–936.

93. Shiokawa S, Nishimura J, Ohshima K, Uike N, Yamamoto K. Establishment of a novel B cell clonality analysis using single-strand conformation polymorphism of immunoglobulin light chain messenger signals. *American Journal of Pathology* 1998;153:1393–1400.

94. Zehentner B, de Baca ME, Wells DA, Loken MR. Intraclonal heterogeneity in concomitant monoclonal lymphocyte and plasma cell populations: combining flow cytometric cell sorting with molecular monoclonality profiling. *Clinical Lymphoma, Myeloma & Leukemia* 2013;13:214–217.

95. Mahe E, Pugh T, Kamel-Reid S. T cell clonality assessment: past, present and future. *Journal of Clinical Pathology* 2018;71: 195–200.

96. Szczepanski T, Beishuizen A, Pongers-Willemse MJ, Hahlen K, Van Wering ER, Wijkhuijs AJ, Tibbe GJ, De Bruijn MA, Van Dongen JJ. Cross-lineage T cell receptor gene rearrangements occur in more than ninety percent of childhood precursor-B acute lymphoblastic leukemias: alternative PCR targets for detection of minimal residual disease. *Leukemia* 1999;13:195–205.

97. Lawnicki L, Rubocki RJ, Chan WC, Lytle DM, Greiner TC. The distribution of gene segments in T-cell receptor gene rearrangements demonstrates the need for multiple primer sets. *Journal of Molecular Diagnostics* 2003;5:82–87.

98. Afzal S, Gil-Farina I, Gabriel R, Ahmad S, von Kalle C, Schmidt M, Fronza R. Systematic comparative study of computational methods for T-cell receptor sequencing data analysis. *Briefings in Bioinformatics* 2017. doi:10.1093/bib/bbx111

99. Lee S, Berg KD, Racke FK, Griffin CA, Eshleman JR. Pseudo-spikes are common in histologically benign lymphoid tissues. *Journal of Molecular Diagnostics* 2000;2:145–152.

100. Murphy K, Berg KD, Geiger T, Hafez M, Flickinger KA, Cooper L, Pearson P, Eshleman JR. Capillary electrophoresis artifact due to eosin: implications for the interpretation of molecular diagnostic assays. *Journal of Molecular Diagnostics* 2005;7:143–148.

101. Albinger-Hegyi A, Hochreutener B, Abdou M-T, Hegyi I, Dours-Zimmermann MT, Kurrer MO, Heitz PU, Zimmermann DR. High frequency of t(14;18)-translocation breakpoints outside of major breakpoint and minor cluster regions in follicular lymphomas: improved polymerase chain reaction protocols for their detection. *American Journal of Pathology* 2002;160: 823–832.

102. Javadekar S, Yadav R, Raghavan SC. DNA structural basis for fragility at peak III of *BCL2* major breakpoint region associated with t(14;18) translocation. *Biochimica Biophysica Acta* 2017;1862:649–659.

103. Hsi E, Tubbs RR, Lovell MA, Braziel RM, Gulley ML. Detection of bcl-2/J(H) translocation by polymerase chain reaction: a summary of the experience of the Molecular Oncology Survey of the College of American Pathologists. *Archives of Pathology & Laboratory Medicine* 2002;126:902–908.

104. McGregor D, Keever-Taylor CA, Bredeson C, Schur B, Vesole DH, Logan B, Chang CC. The implication of follicular lymphoma patients receiving allogeneic stem cell transplantation from donors carrying t(14;18)-positive cells. *Bone Marrow Transplant* 2005;35:1049–1054.

105. Harada S, Sizzle E, Lin MT, Gocke CD. Detection of chromosomal translocation in hematologic malignancies by a novel DNA-based looped ligation assay (LOLA). *Clinical Chemistry* 2017;63:1278–1287.

106. Soldini D, Valera A, Solé C, Palomero J, Amador V, Martin-Subero JI, Ribera-Cortada I, Royo C, Salaverria I, Beà S, Gonzalvo E, Johannesson H, Herrera M, Colomo L, Martinez A, Campo E. Assessment of SOX11 expression in routine lymphoma tissue sections: characterization of new monoclonal antibodies for diagnosis of mantle cell lymphoma. *American Journal of Surgical Pathology* 2014;38:86–93.

107. Yin C, Luthra R. Molecular detection of t(11;14)(q13;q32) in mantle cell lymphoma. *Methods in Molecular Biology* 2013;999: 211–216.

108. Böttcher S, Ritgen M, Buske S, Gesk S, Klapper W, Hoster E, Hiddemann W, Unterhalt M, Dreyling M, Siebert R, Kneba M, Pott C; EU MCL MRD Group. Minimal residual disease detection in mantle cell lymphoma: methods and significance of four-color flow cytometry compared to consensus IGH-polymerase chain reaction at initial staging and for follow-up examinations. *Haemaatologica* 2008;93:551–559.

109. Wijers-Koster P, Droese J, Delabesse E, Spaargaren M, Hermosin L, Diss TC, Mills KI, Jennings BA, Milner BJ, Bloxham D, Al Saati TM, Bassegio L, Kluin PhM, Schuuring E. t(11;14) with BCL1-IGH rearrangement. *Leukemia* 2003;17: 2296–2298.

110. Jiang S, Zhou H, Liang J, Gerdt C, Wang C, Ke L, Schmidt SCS, Narita Y, Ma Y, Wang S, Colson T, Gewurz B, Li G, Kieff E, Zhao B. The Epstein-Barr virus regulome in lymphoblastoid cells. *Cell, Host & Microbe* 2017;22:561–573.

111. La Starza R, Borga C, Barba G, Pierini V, Schwab C, Matteucci C, Lema Fernandez AG, Leszl A, Cazzaniga G, Chiaretti S, Basso G, Harrison CJ, Te Kronnie G, Mecucci C. Genetic profile of T-cell acute lymphoblastic leukemias with *MYC* translocations. *Blood* 2014;124:3577–3582.

112. Nwanze J, Siddiqui MT, Stevens KA, Saxe D, Cohen C. *MYC* immunohistochemistry predicts *MYC* rearrangements by FISH. *Frontiers in Oncology* 2017;7:209.

113. Maurer J, Janssen JW, Thiel E, van Denderen J, Ludwig WD, Aydemir U, Heinze B, Fonatsch C, Harbott J, Reiter A. Detection of chimeric *BCR-ABL* genes in acute lymphoblastic leukaemia by the polymerase chain reaction. *Lancet* 1991;337:1055–1058.

114. Gleissner B, Rieder H, Thiel E, Fonatsch C, Janssen LA, Heinze B, Janssen JW, Schoch C, Goekbuget N, Maurer J, Hoelzer D, Bartram CR. Prospective *BCR-ABL* analysis by polymerase chain reaction (RT-PCR) in adult acute B-lineage lymphoblastic leukemia: reliability of RT-nested-PCR and comparison to cytogenetic data. *Leukemia* 2001;15:1834–1840.

115. van der Velden V, Boeckx N, Gonzalez M, Malec M, Barbany G, Lion T, Gottardi E, Pallisgaard N, Beillard E, Hop WCJ, Hoogeveen PG, Gabert J, van Dongen JJM. Differential stability of control gene and fusion gene transcripts over time may hamper accurate quantification of minimal residual disease—a study within the Europe Against Cancer Program. *Leukemia* 2004;18:884–886.

116. Lion T, Gaiger A, Henn T, Harth E, Haas OA, Geissler K, Gadner H. Use of quantitative polymerase chain reaction to monitor residual disease in chronic myelogenous leukemia during treatment with interferon. *Leukemia* 1995;9:1353–1360.

117. Branford S, Hughes TP, Rudzki Z. Monitoring chronic leukemia therapy by real-time quantitative PCR in blood is a reliable alternative to bone marrow cytogenetics. *British Journal of Haematology* 1999;107:587–599.

118. Muller M, Hardt T, Paschka P, Merx K, La Rosee P, Hehlmann R, Hochhaus A. Standardization of preanalytical factors for minimal residual disease analysis in chronic myelogenous leukemia. *Acta Haematologica* 2004;112:30–33.

119. Moravcova J, Regner J, Mouckova D, Fiser K, Zmekova V, Malacova R, Michalova K, Klamova H. Disease status in patients with chronic myeloid leukemia is better characterized by *BCR-ABL/BCR* transcript ratio than by *BCR-ABL* transcript level, which may suggest a role of normal *BCR* gene in the disease pathogenesis. *Neoplasma* 2005;52:119–125.

120. Hughes T, Kaeda J, Branford S, Rudzki Z, Hochhaus A, Hensley ML, Gathmann I, Bolton AE, van Hoomissen IC, Goldman JM, Radich JP; International Randomised Study of Interferon versus STI571 (IRIS) Study Group. Frequency of major molecular responses to imatinib or interferon alfa plus cytarabine in newly diagnosed chronic myeloid leukemia. *New England Journal of Medicine* 2003;349:1423–1432.

121. Cortes J, Talpaz M, O'Brien S, Jones D, Luthra R, Shan J, Giles F, Faderl S, Verstovsek S, Garcia-Manero G, Rios MB, Kantarjian H. Molecular responses in patients with chronic myelogenous leukemia in chronic phase treated with imatinib mesylate. *Clinical Cancer Research* 2005;11:3425–3432.

122. Wang L, Pearson K, Pillitteri L, Ferguson JE, Clark RE. Serial monitoring of *BCR-ABL* by peripheral blood real-time polymerase chain reaction predicts the marrow cytogenetic response to imatinib mesylate in chronic myeloid leukaemia. *British Journal of Haematology* 2002;118:771–777.

123. White H, Hedges J, Bendit I, Branford S, Colomer D, Hochhaus A, Hughes T, Kamel-Reid S, Kim DW, Modur V, Müller MC, Pagnano KB, Pane F, Radich J, Cross NC, Labourier E. Establishment and validation of analytical reference panels for the standardization of quantitative *BCR-ABL1* measurements on the International Scale. *Clinical Chemistry* 2013;59:938–948.

124. Press R, Kamel-Reid S, Ang D. BCR-ABL1 RT-qPCR for monitoring the molecular response to tyrosine kinase inhibitors in chronic myeloid leukemia. *Journal of Molecular Diagnostics* 2013;15:565–576.

125. Griffiths M, Patton SJ, Grossi A, Clark J, Paz MF, Labourier E; Labceutics International *BCR-ABL1* Standardization Study Group. Conversion, correction, and International Scale standardization: results from a multicenter external quality assessment study for *BCR-ABL1* testing. *Archives of Pathology & Laboratory Medicine* 2015;139:522–529.

126. Seo S, Lee SJ, Park S, Kim MJ, Song JY, Ra EK, Cho SI, Kim HK, Yang MG, Kim JY, Park SS, Seong MW. Performance of two commercially available BCR-ABL1 quantification assays that use an international reporting scale. *Clinical Chemistry and Laboratory Medicine* 2016;54:1157–1160.

127. Slack J, Bi WL, Livak KJ, Beaubier N, Yu M, Clark M, Kim SH, Gallagher RE, Willman CL. Pre-clinical validation of a novel, highly sensitive assay to detect PML-RARa mRNA using real-time reverse-transcription polymerase chain reaction. *Journal of Molecular Diagnostics* 2005;3:141–149.

128. Vizmanos J, Larráyoz MJ, Odero MD, Lasa R, González M, Novo FJ, Calasanz MJ. Two new molecular PML-RARa variants: implications for the molecular diagnosis of APL. *Haematologica* 2002;87:ELT37.

129. de Thé H, Pandolfi PP, Chen Z. Acute Promyelocytic leukemia: a paradigm for oncoprotein-targeted cure. *Cancer Cell* 2017;32:552–560.

130. Mills K, Gilkes AF, Walsh V, Sweeney M, Gale R. Rapid and sensitive detection of internal tandem duplication and activating loop mutations of FLT3. *British Journal of Haematology* 2005;130:203–208.

131. Scholl S, Krause C, Loncarevic IF, Müller R, Kunert C, Wedding U, Sayer HG, Clement JH, Hoffken K. Specific detection of Flt3 point mutations by highly sensitive real-time polymerase chain reaction in acute myeloid leukemia. *Journal of Laboratory and Clinical Medicine* 2005;145:295–304.

132. Estey E. Acute myeloid leukemia: 2013 update on risk-stratification and management. *American Journal of Hematology* 2013;88:318–327.

133. Nassereddine S, Lap CJ, Haroun F, Tabbara I. The role of mutant IDH1 and IDH2 inhibitors in the treatment of acute myeloid leukemia. *Annals of Hematology* 2017;96:1983–1991.

134. Zaneveld J, Hamady M, Sueoka N, Knight R. CodonExplorer: an interactive online database for the analysis of codon usage and sequence composition. *Methods in Molecular Biology* 2009;537:207–232.

135. Kambas K, Mitroulis I, Kourtzelis I, Chrysanthopoulou A, Speletas M, Ritis K. Fast and reliable mutation detection of the complete exon 11-15 *JAK2* coding region using non-isotopic RNase cleavage assay (NIRCA). *European Journal of Haematology* 2009;83:215–219.

Chapter 14

DNA-Based Tissue Typing

Outline

THE MHC LOCUS
HLA POLYMORPHISMS
  HLA Nomenclature
MOLECULAR ANALYSIS OF THE MHC
  Serological Analysis
    HLA Typing
    Screening
    Crossmatching
    Mixed-Leukocyte Reaction
    Protein Gel Electrophoresis
  DNA-Based Typing
    Sequence-Specific Oligonucleotide Probe Hybridization
    Sequence-Specific PCR
    Sequence-Based Typing
    Other DNA-Based Methods
  Combining Typing Results
  HLA Test Discrepancies
  Coordination of HLA Test Methods
ADDITIONAL RECOGNITION FACTORS
  Minor Histocompatibility Antigens
  Nonconventional MHC Antigens
  Killer Cell Immunoglobulin-Like Receptors
MHC DISEASE ASSOCIATION
SUMMARY OF LABORATORY TESTING

Objectives

14.1 Describe the structure and function of the major histocompatibility (MHC) locus.
14.2 Use HLA nomenclature to identify alleles.
14.3 List the human leukocyte antigens (HLAs) that are encoded by the MHC locus, and explain their role in tissue engraftment and rejection.
14.4 Compare and contrast the levels of typing resolution that are achieved by different laboratory methods.
14.5 Describe the laboratory methods used to identify HLAs by serology testing.
14.6 Describe the DNA-based testing methods used for the identification of HLAs.
14.7 Explain how combining different test methods to identify HLAs increases resolution and resolves ambiguities.
14.8 Relate the use of HLA typing for confirming disease diagnosis and predisposition.

The **major histocompatibility complex (MHC)** is a group of genes located on the short arm of chromosome 6. In humans, the MHC gene products are called **human leukocyte antigens (HLAs).** The HLAs were named for their role in the rejection of transplanted organs. Kidney, heart, liver, lungs, skin, pancreas, corneas, blood, bone marrow, and hematopoietic stem cells can be transplanted from one human to another. Transplanted organs, except in the case of identical twins, are **allografts,** indicating genetic differences between the donor of the organ and the recipient. When a transplant is performed, compatibility (matching) of the HLA of the organ donor and the recipient increases the chance for a successful engraftment that will function for several years. If the donor and recipient are not HLA-matched, then the recipient's immune system (primarily mediated by T lymphocytes as well as B lymphocytes) will recognize the donor organ as nonself (foreign) and will mount an immune response against the organ, resulting in its destruction, loss of function, and rejection by the recipient.

The role of the clinical laboratory is to evaluate the HLAs of potential donors and recipients and to aid in the prediction of successful engraftment and the avoidance of graft rejection and graft-versus-host disease (GVHD). Graft rejection results in the failure of the donor organ due to an immune reaction of the recipient against it. GVHD is the reciprocal of graft rejection whereby immunocompetent cells in the donor organ recognize recipient cells as foreign and attack and destroy the recipient cells, resulting in significant morbidity and potential mortality to the recipient. The process of HLA identification utilizes methods targeting cell-associated antigens as well as serum antibodies to "nonself" cells and tissues. Most **tissue typing,** as HLA identification is commonly called, was performed initially by serological methods using antibodies to the different HLAs found in the human population. Increasingly, however, DNA typing methods are being implemented for this purpose, increasing the sensitivity and specificity of the typing procedure.

THE MHC LOCUS

The human MHC locus was discovered in the early 1950s. Investigators independently noted that blood from women who had borne children or from previously transfused persons contained antibodies that agglutinated leukocytes. This discovery led to serological typing methods that originally identified two polymorphic gene loci, HLA-A and HLA-B, followed soon after by the identification of HLA-C and other genes. A test for typing these loci was designed from the observation that large immature mononuclear cells proliferated if lymphocytes from unrelated individuals were mixed and cultured together.[1,2] The results from this **mixed lymphocyte culture (MLC)** reaction did not always agree with the results of serological typing, however. The discrepancy was partly resolved by the discovery of additional genes comprising the HLA-D locus.[3,4]

Historical Highlights

The genetic contribution to transplant rejection was first proposed in 1927 when Bover observed that skin transplants between identical twins were not rejected like those from genetically distinct individuals.[5] The genes involved were first described in mice by Gorer.[6] Snell[7] used mouse cell lines to further define a genetic locus, which he called *H* for *histocompatibility*. Gorer referred to the gene products of this locus as antigen II, and the combined term *H-2* was subsequently used for the MHC locus in mice.

HLAs are divided into three classes (I, II, and III), all encoded by a gene complex located on chromosome 6p (Fig. 14.1). The MHC locus includes genes other than those that code for the HLA. Cytokine genes and genes encoding tumor necrosis factor α (TNF-α) and tumor necrosis factor β (TNF-β) are located inside of the main HLA complex.[8]

In addition to the main MHC locus, gene regions extending beyond the HLA-DP genes toward the centromere and HLA-F toward the telomere comprise the **extended MHC locus (xMHC).** The xMHC locus covers 8 Mb and includes the hemochromatosis gene HLA-F (also called *HFE*), the farthest telomeric gene in the complex.[9,10] The most centromeric locus of the extended MHC is the tapasin region. Tapasin is required for antigenic peptide processing.[11] Some genes that are associated with disease conditions, such as *HFE* linked to the MHC locus, are the basis for the association of

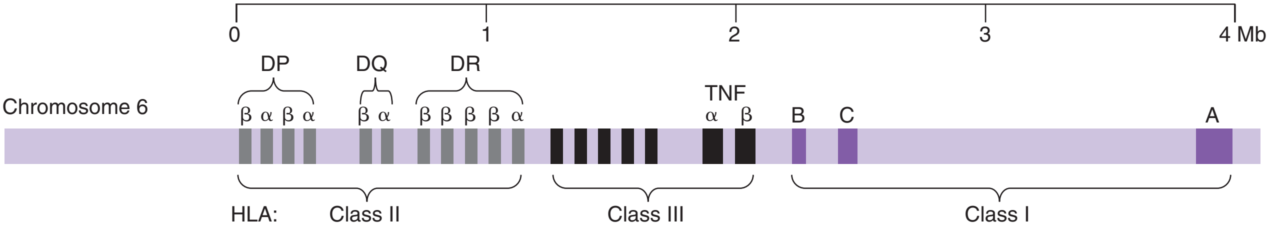


FIGURE 14.1 The MHC locus on chromosome 6 covers about 4 Mb of DNA, depending on the individual. Class I genes are 3 to 6 kb long, and class II genes are 4 to 11 kb in length. TNF-α and TNF-β are not part of the polymorphic HLA system.

TABLE 14.1 Genes of the Major Histocompatibility Locus

MHC Region	Gene Products	Tissue Location	Function
Class I	HLA-A, HLA-B, HLA-C	All nucleated cells	Identification and destruction of abnormal or infected cells by cytotoxic T cells
Class II	HLA-D	B lymphocytes, monocytes, macrophages, dendritic cells, activated T cells, activated endothelial cells, skin (Langerhans' cells)	Identification of foreign antigen by helper T cells
Class III	Complement C2, C4, B	Plasma proteins	Defense against extracellular pathogens
Cytokine genes	TNF-α, TNF-β	Plasma proteins	Cell growth and differentiation

particular disease states with HLA type.[12] The role of the immune system in autoimmune diseases and susceptibility to infections also links HLA type to disease.

Despite the discovery of the MHC gene products as mediators of transplant rejection, recognition of genetically different ("nonself") organs and tissues is not the main function of these glycoproteins. HLAs appear on the surface of cells of the immune system, allowing cell–cell communication during immune functions. Immune reactions involve and are restricted by interactions between T lymphocytes (cells involved in cell-mediated immunity), B lymphocytes (antibody-producing cells), and the MHC molecules.

The gene products of the MHC, class I, II, and III proteins, are present in different amounts on different tissues (Table 14.1). Class I and II are the strongest antigens expressed on cells. Class I molecules (designated as A, B, or C) are expressed on all nucleated cells, whereas class II molecules (designated as D) are only expressed constitutively on "professional antigen-presenting cells," such as B lymphocytes, dendritic cells, and macrophages. As illustrated in Figure 14.2, class I molecules consist of a long (heavy) chain of 346 amino acids (44 kD) associated with a 12-kD peptide, β-2 microglobulin, containing 99 amino acids that are not encoded in the MHC. These two chains are associated with one another on the cell surface by noncovalent bonds. The class I heavy chain displays short branched-chained sugars, making this molecule a glycoprotein. The heavy chain is also a transmembrane polypeptide, anchoring the complex at the surface of the cell. Class II molecules consist of two transmembrane polypeptides, an α chain with three domains, α_1, α_2, and α_3, and a β chain with two domains, β_1 and β_2. The two polypeptides associate, forming a groove between the α_1 and β_1 domains that will hold fragments of antigen that have been engulfed and processed by the cell (extracellular antigens). In contrast, antigens bound to class I molecules (where the peptide-binding domain is formed between the α_1 and α_2 domains) are generated from the processing of macromolecules synthesized within the cell (intracellular antigens).

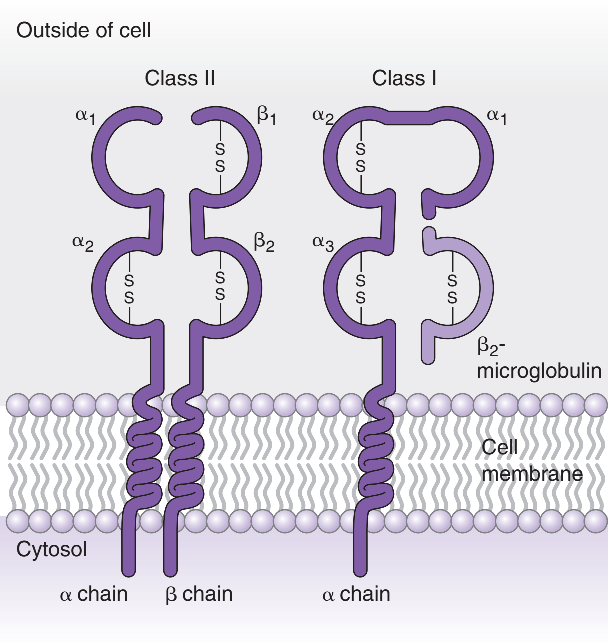


FIGURE 14.2 Class II *(left)* and class I *(right)* polypeptides. Class II antigens consist of two chains, α and β. Class I antigens consist of a heavy chain and a light chain associated together with a molecule of β-2 microglobulin.

Advanced Concepts

Class I and II molecules present fragments of antigens, usually about nine amino acids long, to T lymphocytes. Class I and II molecules vary from one another (are polymorphic), sometimes by a single amino acid. Due to these polymorphisms, different HLA molecules (HLA types) vary in their efficiency of binding antigen fragments, resulting in a range of immune responses to a given antigen. This distinction can affect the symptoms of disease, for example, the likelihood of persons of a particular HLA type infected with HIV to develop full immunodeficiency.[13,14]

HLA POLYMORPHISMS

Genes of the MHC are the most polymorphic genes of the human genome. Polymorphisms in this locus were first defined phenotypically by acceptance or rejection of tissue or by reaction with defined antibodies (serological typing). Molecular typing methods reveal HLA polymorphisms as base changes in the DNA sequence (Fig. 14.3). The changes range from a single base pair (single nucleotide polymorphisms) to loss or gain of entire genes. A particular sequence, or version, of an HLA gene is an **allele** of that gene. The **HLA type** is the collection of alleles detected by phenotypic or genotypic typing methods.

Thus, each HLA gene can differ in sequence from any individual to another, except for identical twins. A set of particular alleles on the same chromosome is a **haplotype** (Fig. 14.4). These alleles are inherited together as a block of chromosomal sequence, unless a rare recombination event within the region separates the alleles. An HLA haplotype is, therefore, the combination of polymorphic sequences or alleles in the HLA gene regions.

Polymorphisms are concentrated in exons 2 and 3 of the class I genes and in exon 2 of the class II genes. These exons code for the amino acids that interact with antigenic peptides, affecting the recognition of nonself peptides. Molecular methods target these exons in HLA-A, HLA-B, and HLA-C class I genes and mostly HLA-DRB class II genes. Other areas including introns have been investigated for potentially useful alleles.[15]

HLA Nomenclature

Polymorphisms are alterations in DNA and/or protein sequences shared by 1–2% of a defined population. Formally, alterations present at lower frequencies are called mutations or variants. Structurally, mutations and polymorphisms are the same thing, changes from a consensus amino acid or nucleotide sequence. Alleles are the different versions of the gene. The polymorphic nature of the MHC, therefore, means that there are multiple alleles of each HLA gene present in the human population. These alleles differ by nucleotide sequence at the DNA level (polymorphisms) and by amino acid sequence and antigenicity at the protein level. Polymorphisms arise mostly as a result of gene-conversion events and rare chromosomal recombinations. Each person will have a

```
  CGG GCC GCG GTG GAC ACC TAC TGC AGA CAC AAC TAC GGG GTT GGT GAG AGC TTC ACA
  CGG GCC GCG GTG GAC ACC TAT TGC AGA CAC AAC TAC GGG GCT GTG GAG AGC TTC ACA
A CGG GCC GCC GTG GAC ACC TAT TGC AGA CAC AAC TAC GGG GCT GTG GNN NNN NNN NNN
  CGG GCC GCG GTG GAC ACC TAC TGC AGA CAC AAC TAC GGG GTT GGT GAG AGC TTC ACA
  --- --- --- --- --- --- --T --- --- --- --- --- --- -C- -TG --- --- --- ---
B --- --- --C --- --- --- --T --- --- --- --- --- --- -C- -TG -** *** *** ***
```

FIGURE 14.3 (A) DNA polymorphisms in the HLA-DRB1 gene. The initial sequence (DRB1*01:01) is written at the top of the panel. Aligned underneath are two alleles of this region, DRB1*01:02 and DRB1*01:03. The bases in color are those that differ from DRB1*01:01. *N* indicates unknown or unsequenced bases. (B) A different way of presenting the alleles. A dash indicates identity to the consensus sequence. Only the polymorphic bases are written. The asterisk indicates unknown or unsequenced bases.

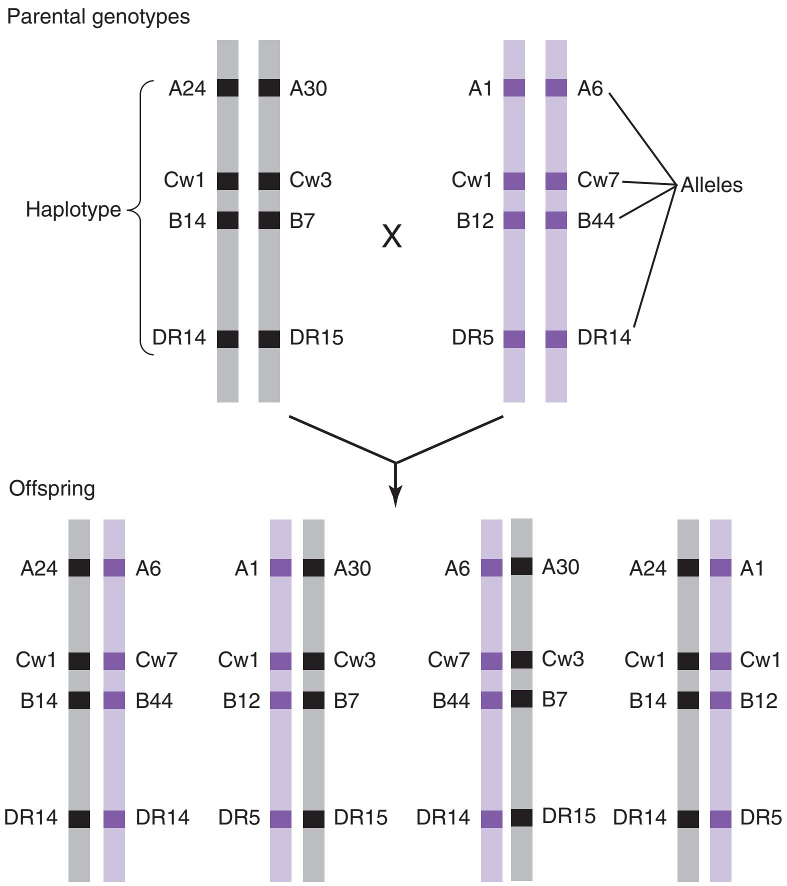

Parental genotypes
Haplotype
A24
Cw1
B14
DR14
A30
Cw3
B7
DR15
X
A1
Cw1
B12
DR5
A6
Cw7
B44
DR14
Alleles
Offspring
A24
Cw1
B14
DR14
A6
Cw7
B44
DR14
A1
Cw1
B12
DR5
A30
Cw3
B7
DR15
A6
Cw7
B44
DR14
A30
Cw3
B7
DR15
A24
Cw1
B14
DR14
A1
Cw1
B12
DR5

FIGURE 14.4 A haplotype is the combination of alleles that are inherited together. In this example, parental genotypes *(top)* can produce four possible genotypes in the offspring *(bottom)*.

particular group of HLA alleles inherited from his or her parents. The maternal and paternal HLA antigens are expressed codominantly on cells.

HLA alleles were first defined at the protein level by antibody recognition (serologically). A standard nomenclature for expressing serologically defined antigens was established by the World Health Organization (WHO) Nomenclature Committee for Factors of the HLA System. In this system, *HLA* refers to the entire gene region, and *A, B,* and *D* refer to the particular loci; for example, HLA-A, HLA-B, or HLA-D. The HLA-D locus consists of subregions P, Q, M, O, and R, termed HLA-DP, HLA-DQ, HLA-DM, and so forth. Each of these subregions consists of genes that code for either an α- or β-chain polypeptide; for example, the first β polypeptide is encoded in the HLA-DRB1 gene (see Fig. 14.1). A small *w* is included in HLA-Cw, HLA-Bw4, and HLA-Bw6 allele nomenclature. The *w* denotation was originally a designation of alleles in "workshop" status or found in high prevalence in the population. Workshop designation is no longer required for the HLA-C locus; however, the *w* was retained for HLA-C alleles to distinguish them from the C designation used for complement genes. The *w* is retained with HLA-Bw4 and HLA-Bw6, which are considered "public" (high-prevalence) antigens.

A list of the serologically defined alleles of the HLA genes accepted by the WHO is shown in Table 14.2. The WHO official nomenclature refers to serologically defined alleles: a number follows the gene region name; for example, HLA-B51 denotes HLA-B antigen 51 (defined by reaction to a known antibody). Number designations of new alleles of a previously defined allele with broad specificity (parent allele) are followed by the number of the parent allele in parentheses. For example, HLA-A24(9) denotes the HLA-A antigen 24 from parent antigen 9. The derived antigens are called **split specificities.** Additional antigens have been defined by reactions between known antigens and serum antibodies (antiserum reactivity). The number of HLA alleles increase as new specificities are defined.

With the introduction of molecular biology techniques in the 1980s, HLA typing at the DNA level required nomenclature for specific DNA sequences.[16,17] Many new alleles have been and are currently being defined at the DNA level. A revised nomenclature is used for denoting alleles defined by DNA sequence. The alleles are named in sequential order as they are discovered. The gene name, such as HLA-DRB1, is followed by an asterisk or separator (*), the allele sequence family number (allele group or type), and then a number for the specific allele (DNA sequence) separated by a colon or field separator. For example, A*25:03:07 is the seventh polymorphic HLA protein of the third specific allele, 03, of the HLA-A*25 family of alleles. Additional information not routinely used in identifying alleles is also part of their nomenclature. Silent mutations (changes in the DNA sequence that do not change the amino

TABLE 14.2 Serologically Defined HLA Specificities*

HLA-A	HLA-B	HLA-C	HLA-DR	HLA-DQ	HLA-DP
A1	B5	Cw1	DR1, DR103	DQ1	DPw1
A2, A203, A210	B51(5), B5102, B5103	Cw2	DR2	DQ5(1)	DPw2
A3	B52(5)	Cw3	DR15(2)	DQ6(1)	DPw3
A9	B7, B703	Cw9(w3)	DR15(2)	DQ2	DPw4
A23(9)	B8	Cw10(w3)	DR3	DQ3	DPw5
A24 (9), A2403	B12	Cw4	DR17(3)	DQ7(3)	DPw6
A10	B44(12)	Cw5	DR18(3)	DQ8(3)	
A25(10)	B45(12)	Cw6	DR4	DQ9(3)	

TABLE 14.2 Serologically Defined HLA Specificities *(Continued)*

HLA-A	HLA-B	HLA-C	HLA-DR	HLA-DQ	HLA-DP
A26(10)	B13	Cw7	DR5	DQ4	
A34(10)	B14	Cw8	DR11(5)	DQ3	
A66(10)	B64(14)	Cw10(w3)	DR12(5)		
A11	B65(14)		DR6		
A19	B15		DR13(6)		
A74(19)	B62(15)		DR14(6), DR1403, DR1404		
A68(28)	B63(15)		DR7		
A69(28)	B75(15)		DR8		
A29(19)	B76(15)		DR9		
A30(19)	B77(15)		DR10		
A31(19)	B16		DR51		
A32(19)	B38(16)		DR52		
A33(19)	B39(16), B3901, B3902		DR53		
A36	B17		DR16(2)		
A43	B57(17)				
A80	B58(17)				
A28	B18				
A74 (19)	B21				
	B49(21)				
	B50(21)				
	B22				
	B54(22)				
	B55(22)				
	B56(22)				
	B27, B2708				
	B35				
	B37				

Continued on following page

TABLE 14.2 Serologically Defined HLA Specificities *(Continued)*

HLA-A	HLA-B	HLA-C	HLA-DR	HLA-DQ	HLA-DP
	B40, B4005				
	B60(40)				
	B61(40)				
	B41				
	B42				
	B46				
	B47				
	B48				
	B53				
	B59				
	B67				
	B70				
	B71(70)				
	B72(70)				
	B73				
	B7801				
	B81				
	Bw4				
	Bw6				
	B51(5), B51:02, B51:03				
	B52(5)				
	B57(17)				
	B58(17)				
	B62(15)				
	B63(15)				
	B64(15)				
	B65(15)				

*Alleles are listed as broad antigen groups, followed by split antigens for that group, with the broad antigen number in parentheses. Associated antigens, such as B40 and B4005, are listed together.

acid sequence), also called **synonymous** changes, are designated by a number following the specific allele number. For example, A*03:01:02 indicates a synonymous allele, 02, of the first specific allele, 01, from the HLA-A*03 family of alleles. A fourth number designation indicates that the synonymous change results from a polymorphism in intronic sequences beyond the coding regions (exons) of the genes. Thus, A*26:07:01:01 is the 01 subtype allele of the HLA-A*26:07 allele, the second 01 indicating that the polymorphism is in an adjacent intron.

Optional suffixes can also be added to the allele name. The letter *N* following the specific allele number indicates a **null allele,** or phenotypic absence of that antigen. For example, B*13:07N denotes the 07 allele of the 13-allele family in the HLA-B locus at the DNA level. At the protein level, however, the encoded protein does not react with any antibody. The B*13:07N null allele is due to a 15-bp deletion in the HLA-B gene. Null alleles can also result from nonsense, frameshift, splice site, or other premature stop mutations that prevent translation of the amino acids destined to bind to antibody.

Other descriptive designations are less frequently used, including L, S, Q, A, C, X, and ?. The letters *L* and *S* indicate poor expression of the allele at the cell surface or a soluble allele, respectively. *Q* is used to indicate questionable allele expression based on the effect of the DNA mutation on expression of other alleles. If no other information is available, an *A* indicates aberrant expression without confidence that the allele is actually expressed. *C* indicates allele expression in the cytoplasm, rather than on the cell surface. (*C* and *A* have yet to be used in allele names.) An *X* notation indicates that resolution of the allele was not done, for example, B*08X. A question mark (?) indicates that the resolution was not clear: B*08?.

Advanced Concepts

The National Marrow Donor Program assigns alphabetical allele codes, such as AD or RJH, to allele combinations from submitted requests. Generic codes can be used with several loci and allele families. For example, the combination of alleles, 15:01/15:01N, 15:11/15:15, 15:33/15:34, 15:57/15:60, in any HLA gene is designated RDX so that B*15:01/15:01N, 15:11/15:15, 15:33/15:34, 15:57/15:60 = B*15RDX. Allele-specific codes are used for allele combinations that include more than one serological family or that contain an N, L, or S expression.

Ambiguity is the recognition of two or more antigens by the same antibody, or cross-reaction, so that the exact allele cannot be called. Ambiguity is designated by a slash (/) between the possible allele numbers or an en dash (–) for a series of alleles in which the first and last allele are named. For example, if a typing test results in either B*07:33 or B*07:35, the notation is B*07:33/B*07:35. If a typing test indicates that the allele is either B*07:33, B*07:34, B*07:35, or B*07:36, the designation is B*07:33–B*07:36. Ambiguity also arises from the inability of some typing methods to assign heterozygous alleles to one or the other chromosome. A combination of methods may be used to resolve ambiguities. Family studies are also helpful in this regard.

Resolution is the level of detail to which the allele is determined. Low resolution identifies broad allele types or groups of alleles. A typing of A*26 is low resolution, which can be determined at the serological level. Typing methods that detect specific alleles in addition to the identification of all serological types are at medium resolution. The typing result A*26:01/A*26:05/A*26:10/A*26:15 is medium resolution. High-resolution typing procedures can discriminate between almost all specific alleles. A*26:01 is high resolution determined by DNA analysis. A range of methods, from serological typing to direct DNA sequence analysis, affords the laboratory a choice of low-, medium-, or high-resolution typing. Whereas low resolution is adequate for solid organ transplantation typing, bone marrow or stem cell transplants require high-resolution methods.

MOLECULAR ANALYSIS OF THE MHC

There are over 17,000 HLA and related alleles described in the International ImMunoGeneTics Project (IMGT) IPD-IMGT/HLA Database (Table 14.3). Genetic

TABLE 14.3 Number of HLA Alleles Identified Serologically and by DNA Sequence*

Gene	Serology	Genetic	Gene	Serology	Genetic
Class I			HLA-DQA1		94
HLA-A	28	4,081	HLA-DQB1	9	1,178
HLA-B	29	4,950	HLA-DMA		7
HLA-C	10	3,685	HLA-DMB		13
Class II			HLA-DPA1		64
HLA-DRA		7	HLA-DPB1		963
HLA-DRB1	18	2,146	HLA-DOA		12
HLA-DRB2		1	HLA-DOB		13
HLA-DRB3	1	152	*Extended MHC*		
HLA-DRB4	1	74	HLA-E		6
HLA-DRB5	1	55	MICA		107
HLA-DRB6		3	MICB		42
HLA-DRB7		2	TAP1	6	12
HLA-DRB8		1	TAP2	4	12
HLA-DRB9		6			

*The numbers represent the number of named alleles for each gene as listed by IMGT (http://www.ebi.ac.uk/ipd/imgt/hla). The World Marrow Donor Association Quality Assurance and Working Group on HLA Serology to DNA Equivalents publishes a comprehensive dictionary of antigen and allele equivalents.

(DNA-based) typing concentrated in the HLA-A, HLA-B, HLA-C, and HLA-DRB1 genes has identified alleles that far outnumber the serological alleles defined for these genes. Over 400 new nucleotide sequences were defined in just 2 years, from 2002 to 2004. The WHO Nomenclature Committee devised rules for the submission of new alleles for official numerical designation.[18] Allele sequences are stored in the GenBank, the European Molecular Biology Laboratory, and the DNA Data Bank of Japan databases. A list of newly reported alleles is published monthly in the journals *Tissue Antigens, Human Immunology,* and the *International Journal of Immunogenetics*. A comprehensive dictionary of antigen-DNA sequence allele equivalents is published periodically.[19]

Identification of alleles in the laboratory serves several purposes. In addition to the selection of organ donors, the extent of HLA-type matching between donor and recipient predicts the long-term survival of the donor organ in the recipient. Furthermore, because disease genes are located in and around the MHC locus, certain HLA gene alleles are linked to disease, affording another aid in diagnosis or prediction of disease phenotypes.

There are three approaches to the analysis of HLA alleles in the HLA laboratory: typing, screening, and crossmatching. Typing is the initial identification of the HLA alleles of a specimen through protein or DNA-based methods. Typing may be used both to define HLA haplotypes and to look for specific HLA types that are linked to disease states. Screening is the detection of anti-human antibodies in serum that match known HLA alleles. **Crossmatching** is the more specific screening of recipient sera for antibodies against antigens displayed by potential organ donors.

Serological Analysis

Traditionally, HLA typing for organ transplantation was performed serologically; that is, by antigen–antibody recognition. Although serological testing yields only low-resolution typing results, there are some advantages to these methods. Serological typing is a relatively rapid method that reveals immunologically relevant epitopes. Also, these studies can be used to resolve ambiguities or confirm null alleles detected by other methods. Serological tests include HLA phenotype determination, in which patient cells are tested with known antisera (HLA typing), and screening of patient sera for anti-HLA antibodies.

HLA Typing

Lymphocytes are HLA-typed using the **complement-dependent cytotoxicity (CDC)** test (Fig. 14.5).[20] In this procedure, multiple alleles are determined using a panel of antibodies against known HLA types. These antibodies are prepared from cell lines or from donors with known HLA types. Plates preloaded with antibodies **(typing trays)** are commercially available. Alternatively, some laboratories construct their own antibody panels. The collection of antibodies can be modified to represent ethnic populations or antigens of high prevalence in particular geographical areas. As the antibody preparations are used repeatedly, the antigen binding characteristics of the various antibodies are recognized and recorded. Experienced technologists have detailed documentation of antibody panels, including which antibodies bind antigen well and which antibodies bind less strongly.

To begin the typing procedure, different antibodies are placed in each well of the typing tray. Donor or recipient lymphocytes to be typed are distributed to the wells. Cross reactivity is assessed by the uptake of trypan blue or eosin red dye in cells that have been permeabilized due to reaction with the antibody and with complement that is activated by the antigen–antibody complexes (Fig. 14.6). Cytotoxicity is scored by the estimated percentage of cells in a well that have taken up the dye. The American Society for Histocompatibility and Immunogenetics (ASHI) developed guidelines for the numerical description of the observed cytotoxicity (Table 14.4). High cytotoxicity (reading >6) in a well of the plate indicates that the cells being tested have cell surface antigens matching the known antibody in that well. Because reading is somewhat subjective, it is recommended that trays be read by at least two technologists independently. An example of partial results from

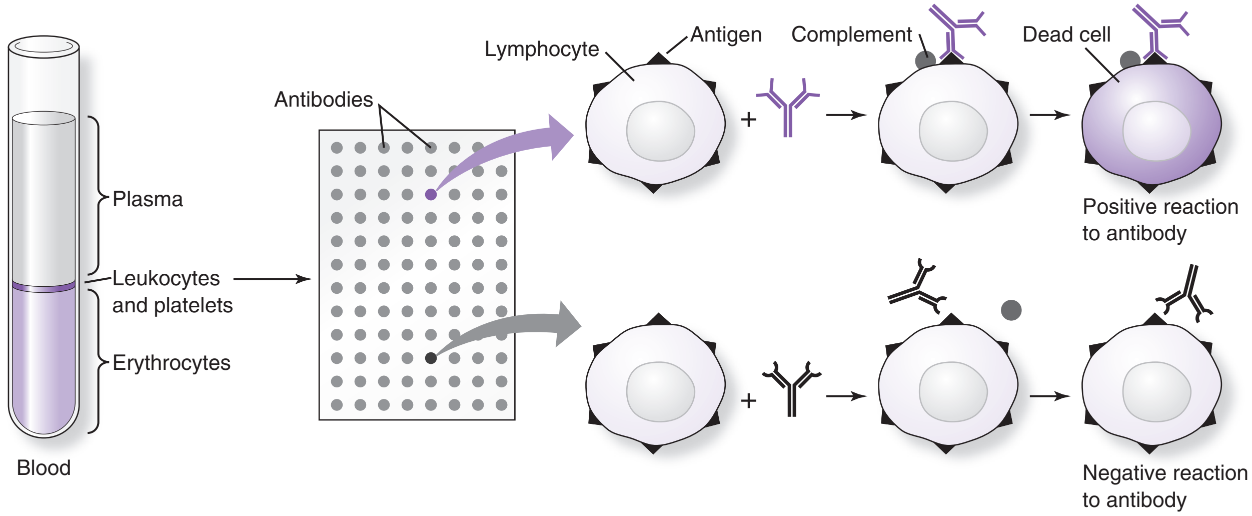


FIGURE 14.5 Crossmatching to known antibodies is performed on lymphocytes *(buffy coat, left)* in a 96-well plate format where each well contains different known antibodies. If the antibody matches the cellular antigen *(positive reaction, top)*, complement-dependent cytotoxicity will occur, and the dead cell will take up stain *(green)*. If the antibody does not match the cellular antigen, there is no cytotoxicity.

TABLE 14.4 Expression of CDC

% Dead or Lysed (dyed)	Interpretation	Score
0–10	Negative	1
11–20	Doubtful negative	2
21–50	Weak positive	4
51–80	Positive	6
81–100	Strong positive	8
	Unreadable	0

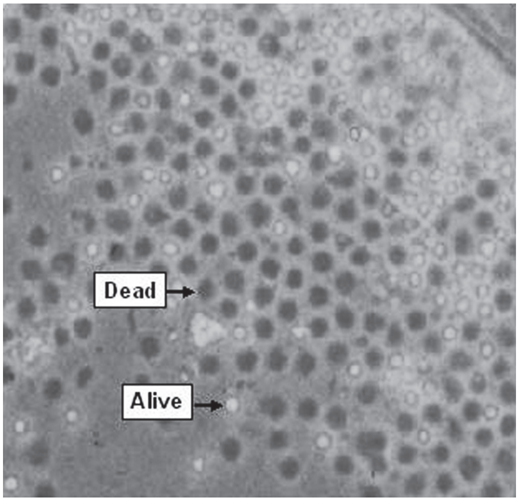


FIGURE 14.6 Cells stained for cytotoxicity. Dead cells take up dye, and live cells remain transparent. *(Photo courtesy of Dr. Andres Jaramillo, Rush University Medical Center.)*

a CDC test is shown in Table 14.5. The results indicate that the tested cells are HLA-type A28, B44.

Each HLA antigen has multiple epitopes, some unique to that HLA antigen and some **cross-reactive epitope groups (CREGs)** shared by other HLA antigens. The CDC test can be performed with **private antibodies,** those that bind to one specific HLA type, or antibodies that bind to CREG (or "public" antigens). Analysis of antibodies with shared specificities aids in narrowing the specificity, as illustrated in Table 14.5. As shown, all antibodies with A28- or B44-specificity generated cytotoxicity, supporting the determination of an A28, B44 haplotype.

TABLE 14.5 Example of Results From a CDC Assay

Antibody	Score*
A2, A28, B7	8
A2, A28	6
A10	1
A10, A11	1
B7, B42	8
B7, B27	8
B7, B55	8
B44, B45, B21	6
B44, B45	8
B44	8
B45	1

*Scores are 1–8, depending on the percentage of dyed cells observed.

CREG matching or residue matching (determined from the amino acid sequences of the antigens) is considered for kidney transplant screening in order to define the spectrum of HLA-specific antibodies more precisely. Some epitopes are more important than others with respect to organ rejection. Therefore, certain mismatches are allowable if the critical epitopes match.

Screening

Successful organ transplant depends on the minimal reaction of the recipient immune system to the antigens of the donor organ. Normal sera do not have antibodies against human antigens, termed **anti-human antibodies** or **alloantibodies.** Persons who have had a previous organ transplant, blood transfusions, or pregnancies, however, will have anti-human antibodies (termed **humoral sensitization**) that may react against a new donor organ. The chance of a successful transplant is improved by defining the specificity of the alloantibodies and selecting a suitable organ that does not have HLA antigens corresponding to the antibodies in the patient's sera.[21]

Humoral sensitization and the identity of alloantibodies present in the recipient serum are determined in a modified version of the CDC assay using the patient's serum as the source of antibodies and reference lymphocytes of known HLA types prevalent in the general population. The reference lymphocytes are defined by their recognition of **panel reactive antibodies (PRAs).** This test against PRA estimates the percentage of the general population with whom the patient will cross-react.

The percentage of the panel of lymphocytes killed by the sera is referred to as %PRA. Patients with %PRA activity of more than 50% are considered to be highly sensitized; finding crossmatch-negative donors is more difficult in these cases.[22]

Screening of sera with microparticles (beads) is performed in laboratories with flow-cytometry capability (Fig. 14.7). For this method, the beads are coupled to pools of antigens derived from cell lines of defined HLA types. The beads are then exposed to test serum, and those beads carrying antigens that are recognized by antibodies present in the test serum will bind to those antibodies. After removal of unbound antibodies, a fluorescently labeled secondary reporter antibody is applied, and the antibody-bound beads are detected by flow cytometry. The advantages of this method over the CDC test are that the reaction is performed in a single tube and there is less subjectivity in the interpretation of results. Because this test uses pooled antigens, however, it can detect the prevalence of anti-human antibodies in the test serum, but it cannot identify which specific antibodies are present. A negative result does preclude further alloantibody assessment. A variation of this method developed commercially utilizes beads with their own internal fluorescence. By conjugating known antigens to beads of different internal fluorescence, the positively reacting antibodies can be specifically identified while still performing the test in the same tube.

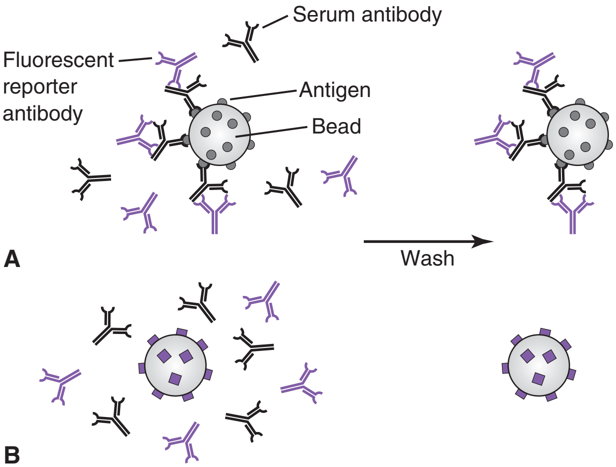


FIGURE 14.7 Detection of serum antibodies using bead arrays. In this illustration, separate preparations of beads are conjugated to two different known antigens. The patient serum tested contains an antibody to the antigen on the beads in (A) but not the antigen on the beads in (B). A secondary antibody targeting the bound serum antibody generates a fluorescent signal detected by flow cytometry. If a matching antibody is not present in the test serum as in (B), no antibody will be bound.

Crossmatching

The CDC test is also used for crossmatching potential organ donors and recipients. For crossmatching, recipient serum is the source of antibodies tested against donor lymphocytes (Fig. 14.8). If the recipient serum kills the donor lymphocytes, it is a positive crossmatch and contraindication for using the crossmatched donor.

Advanced Concepts

More detailed crossmatch information is achieved by separate analysis of donor B and T lymphocytes. Unactivated T cells display class I antigens, and B cells display both class I and class II antigens. Therefore, if B cells cross-react with the serum antibodies and T cells do not, the serum antibodies are likely against class II antigens.

Other methods used for crossmatching include variations on the lymphocytotoxicity assay and nonlymphocytotoxic methods that utilize flow cytometry, such as the bead arrays just described. Alternative methods also include enzyme-linked immunosorbent assay (ELISA) using solubilized HLA antigens. ELISA can be used to monitor the change in antibody production over time or humoral sensitization developing after the transplant.

Mixed-Leukocyte Reaction

T lymphocytes are primarily responsible for cell-mediated organ rejection. The mixed leukocyte culture or

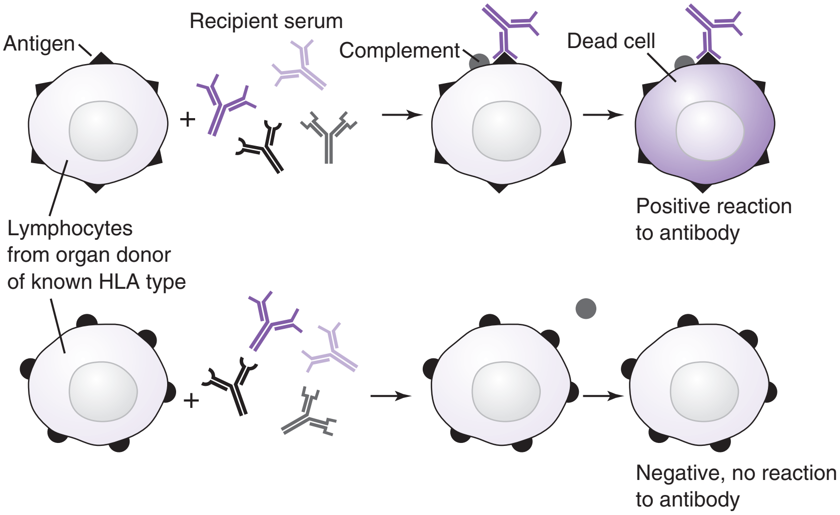

Antigen
Recipient serum
Complement
Dead cell
Lymphocytes from organ donor of known HLA type
Positive reaction to antibody
Negative, no reaction to antibody

FIGURE 14.8 In the crossmatch by CDC assay, the recipient serum is the source of antibodies to type lymphocytes from potential organ donors. The antibodies in the recipient serum can be identified if the HLA type of the lymphocytes is known.

mixed lymphocyte reaction (MLR) is an in vitro method used to determine T-cell cross reactivity between donor and recipient. The MLR assay measures the growth of lymphocytes activated by cross reactivity as an indication of donor–recipient incompatibility. MLR can also test for cell-mediated cytotoxicity and cytokine production, by either the donor or recipient lymphocytes, so it can be used to predict GVHD as well as recipient-mediated transplant rejection. For this test, cells must be incubated together for several days. Cell activation and growth are assessed by uptake of ^{3}H-thymine. Even though the MLR is more likely to detect HLA mismatches than serology techniques, the time and technical demands of the MLR precluded its routine use for pretransplant histocompatibility testing in the clinical laboratory.[23]

Protein Gel Electrophoresis

The protein products of the HLA genes can be distinguished by mobility differences in one-dimensional gel isoelectric focusing or two-dimensional gel electrophoresis methods. In addition to clinical applications, these typing methods were also applied to forensic identification. Protein methods are limited by the demands of the methodology and the ability to distinguish only those proteins that have different net charges. Mass spectrometry has also been applied to the evaluation of HLA peptides for tissue and organ transplantation (immunopeptidomics).[24,25]

DNA-Based Typing

Typing, screening, and crossmatch analysis are critical for the selection of potential donors and successful engraftment. Methods differ in sensitivity. The choice and design of the method used, therefore, will affect the ability to predict rejection risk.

Limitations of serological and protein-based methods have led to the development of more refined DNA typing with higher powers of resolution, especially for bone marrow transplantation. One of the first DNA methods for molecular typing was restriction fragment length polymorphism (RFLP) analysis by Southern blot to identify HLA class II alleles. Studies showed that kidneys matched by RFLP typing survived longer than those matched by serological typing. Just as with other applications of molecular testing, the development of amplification and direct sequencing methods greatly advanced the analysis of HLA polymorphisms at the DNA sequence level.

DNA typing focuses on the most polymorphic loci in the MHC, HLA-B, and HLA-DRB. HLA-A, HLA-B, HLA-C, and HLA-DRB are all considered important for successful transplantation outcomes.[26] For this reason, the number of alleles in these particular loci has risen significantly compared with the number of serologically defined polymorphisms.

Whole-blood patient specimens collected in ethylenediaminetetraacetic acid (EDTA) anticoagulant are used for DNA-based typing. Cell lines of known HLA type are used for reference samples. Standards and quality assurance for DNA-based assays have been established by ASHI (http://www.ashi-hla.org). DNA isolation can be performed from white blood cell preparations (buffy coat) or from isolated nuclei treated with proteinase K.

Sequence-Specific Oligonucleotide Probe Hybridization

Hybridization of a labeled probe to immobilized amplicons of the HLA genes (dot blot) was one of the first methods that utilized polymerase chain reaction (PCR)–amplified DNA for HLA typing. For this procedure (Fig. 14.9), the HLA region under investigation is amplified by PCR using primers flanking the polymorphic sequences. Because the majority of polymorphic sequences are located in exon 2 of the class II genes and exons 2 and 3 of the class I genes, primers are designed to target these regions.

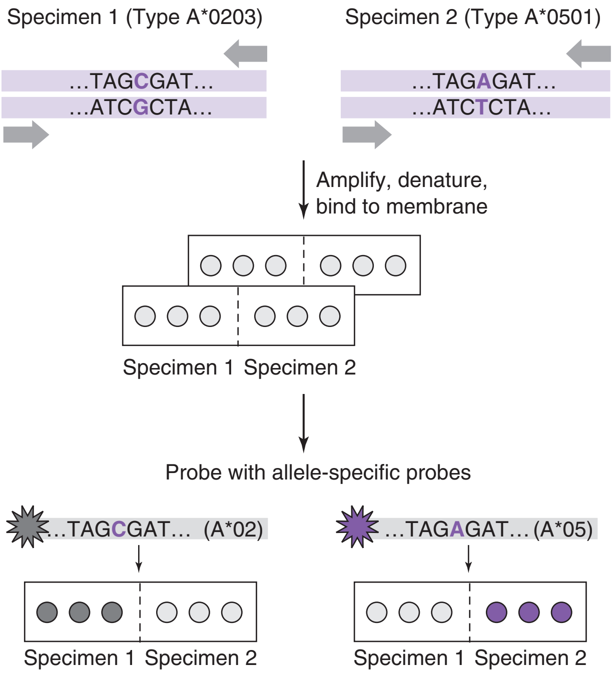


FIGURE 14.9 The principle of the SSOP assay is shown. An HLA gene region is amplified from specimen DNA using generic primers *(top)*. The amplicons are immobilized on a membrane and probed with labeled sequences complementary to specific alleles. Signal from the bound probe will indicate the allele of the immobilized DNA.

An assay, for example, using 30 probes required approximately 70 µl of PCR product. The amplicons were denatured by addition of NaOH and spotted onto a membrane in 1- to 2-µl volumes. Spotting was done manually with a multichannel pipet or by a vacuum manifold with a 96-well plate format. The spotted DNA was dried, then permanently attached to the membrane by ultraviolet (UV) cross-linking (exposure to UV light) or baking. Separate membranes were produced for each probe to be used. Every membrane included reference amplicons complementary (positive control) and non-complementary (negative control) to all probes in the assay. Spotting consistency was checked using a consensus probe that will hybridize to all specimens on a membrane.

The probes used in this type of assay were short (19 to 20 bases), single-stranded DNA chains (oligonucleotides) designed to hybridize to specific HLA alleles. The oligonucleotides were labeled, that is, covalently attached to biotin or digoxygenin. Probe sequences were based on sequence alignments of HLA polymorphic regions, aligned so that the polymorphic nucleotide was in the middle of the probe sequence. Hybridization conditions depended on the optimal hydrogen bonding of probe complementary to a test sequence in comparison with another sequence, differing from (not complementary to) the probe by at least one base. In the assay, spots of immobilized specimens bound to specific probes gave a positive colorimetric or chemiluminescent signal from the labeled probe. Panels of probes defined specific alleles according to which probe bound the immobilized amplified DNA under investigation. The number of probes used depended on the design of the assay. For example, an intermediate resolution assay of the HLA-DRB locus might have taken 30 to 60 probes. Studies achieved high-resolution identification of the majority of HLA-A, HLA-B, and HLA-C alleles using 67 HLA-A, 99 HLA-B, and 57 HLA-C probes and intermediate resolution with 39 HLA-A and 59 HLA-B alleles.[27]

Sequence-specific oligonucleotide probe (SSOP) was also performed in a reverse dot-blot configuration in which the allele-specific probes were immobilized on the membrane (Fig. 14.10). In this method, the specimen DNA was labeled by PCR amplification using primers covalently attached to biotin or digoxygenin at the 5′ end. In contrast to the SSOP described previously where amplicons from each specimen were spotted on multiple membranes, each specimen was tested for multiple alleles on a single membrane. Therefore, instead of having a separate membrane of multiple specimens for each probe, a separate membrane of multiple probes was required for each specimen. The reverse dot-blot strategy is now applied to bead array systems where fluorescently distinct beads carry the oligonucleotide probes. This system is used for typing of HLA-A, HLA-B, and HLA-C class I and HLA-DRB1, DRB3, DRB4, DRB5, and DQB1 antigens.

SSOP is considered low to intermediate resolution, depending on the number and types of probes used in the assay. Because some probes have multiple specificities, hybridization panels can be complex. Computer programs may be used for accurate interpretation of SSOP results.

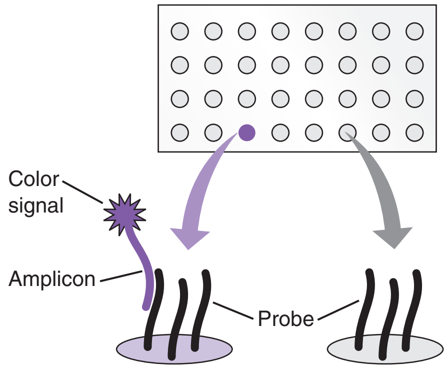

Color signal
Amplicon
Probe

FIGURE 14.10 In reverse dot-blot SSOP, the probe is immobilized on the membrane. Patient DNA is amplified using primers covalently bound to biotin or digoxygenin at the 5+ end. The amplicons are then hybridized to panels of probes immobilized on a membrane *(top)*. If the sequence of the amplicon matches and hybridizes to that of the probe, a secondary reaction with enzyme-conjugated avidin or antidigoxygenin will produce a color or light signal when exposed to substrate. If the sequence of the amplicon differs from that of the probe, no signal is generated *(bottom right)*.

Sequence-Specific PCR

A faster method of sequence-based typing is the use of sequence-specific primers that will amplify only specific alleles (Fig. 14.11). As previously noted, the 3′ end of a PCR primer must be complementary to the template for recognition by DNA polymerase. By designing primers that end on the polymorphic bases, successful generation of a PCR product will occur only if the test sequence has the polymorphic allele complementary to the primer. Detection of the PCR product is used to indicate specific alleles. Sequence-specific PCR (SSP-PCR) is faster and easier than SSOP in that no probes or labeling steps are required, and the results of SSP-PCR are determined directly by agarose gel electrophoresis.

For SSP-PCR, isolated DNA is amplified using sets of primers designed to specifically amplify a panel of alleles. Reactions are set up in a 96-well plate format, with different allele- or sequence-specific primer sets in each well. Each PCR reaction mix contains sequence-specific primers and amplification control primers in a multiplex format. The amplification control primers should yield a product for every specimen (except the

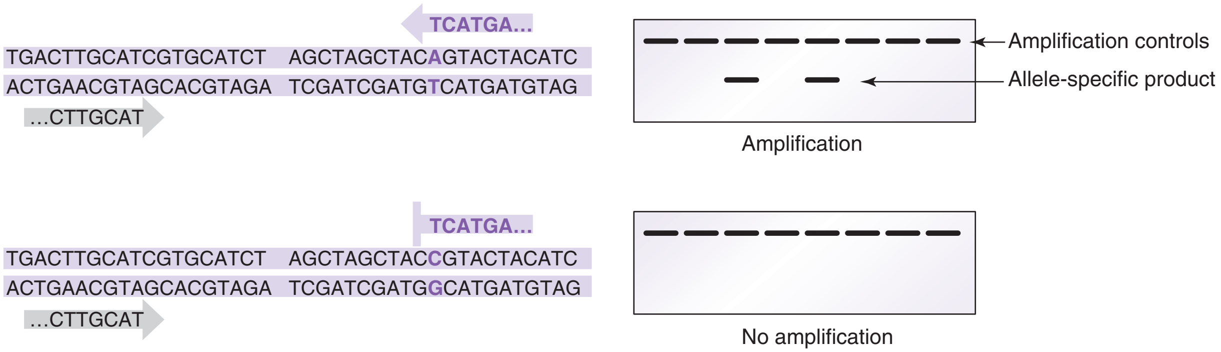

TCATGA...
TGACTTGCATCGTGCATCT AGCTAGCTACAGTACTACATC
ACTGAACGTAGCACGTAGA TCGATCGATGTCATGATGTAG
...CTTGCAT
Amplification controls
Allele-specific product
Amplification
TCATGA...
TGACTTGCATCGTGCATCT AGCTAGCTACCGTACTACATC
ACTGAACGTAGCACGTAGA TCGATCGATGGCATGATGTAG
...CTTGCAT
No amplification

FIGURE 14.11 Sequence-specific PCR relies on the requirement of complementarity between the 3′ base of the primer and the template. The sequence-specific primer ending in AGTACT will be extended only from a template carrying the polymorphism shown.

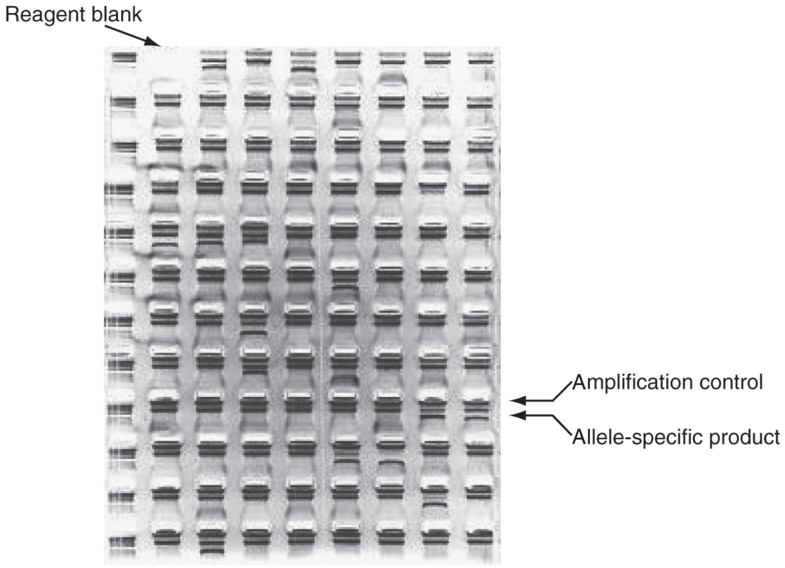


FIGURE 14.12 Results of an SSP-PCR of 95 primer sets detected by agarose gel electrophoresis show the presence of specific alleles as allele-specific PCR products. An amplification control *(top band in each well)* is included with each reaction to avoid false-negative results due to amplification failure. Contamination is monitored by a reagent blank. *(Photo courtesy of Christin Braun, Rush University Medical Center.)*

negative control). The sequence-specific primers should only yield a product if the specimen has the allele complementary to (matching) the allele-specific primer sequence. The amplification primers are designed to yield a PCR product of a size distinct from the product of the allele-specific primers. The two amplicons can then be resolved by agarose gel electrophoresis. An illustration of the results expected from SSP-PCR is shown in Figure 14.12. Specimens will yield two PCR products (amplification control and allele-specific product) only from those wells containing primers matching the specimen HLA allele. Wells containing primers that do not match the patient's HLA allele will have a band only from the amplification control.

SSP-PCR is frequently performed on PCR plates preloaded with reaction mixes (typing trays) containing primers specific for class I and class II DRB and DRQ genes. Specimen DNA is introduced into the individual wells, and the plates are placed in the thermal cycler. Plate maps of allele-specific primers are provided for interpretation of the HLA-type. SSP-PCR has become a commonly used method for testing potential donors before transplant.

Sequence-Based Typing

The most definitive way to analyze DNA at the nucleotide sequence level is by direct DNA sequencing. This is true for any DNA test, no less for discovery and identification of HLA types. Classical **sequence-based typing (SBT)** involves amplification of HLA class I exons 2, 3, and 4 and class II exons 2 and 3 (Fig. 14.13). The larger amplicons are purified and added to a sequencing reaction mix for sequencing with the inner sequencing primers (Fig. 14.14). Following gel or capillary gel electrophoresis, the fragment patterns or electropherograms of the DNA ladders are examined for specific polymorphisms (matches to stored sequence databases of known alleles).

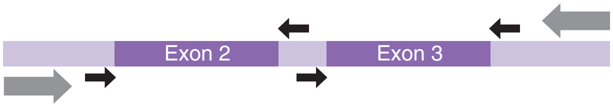


FIGURE 14.13 Example of primer placement for amplification and Sanger sequence analysis. PCR primers *(outer large arrows)* are used to amplify the region of interest. The PCR product is then sequenced using four different primers *(inner small arrows)* in separate sequencing reactions.

There are some technical challenges to sequence-based typing, such as secondary structure and other artifacts that alter the mobility of fragments and complicate interpretation. Furthermore, several events can occur that will compromise sequence quality. Manufacturers of sequencing reagents supply ways to correct most

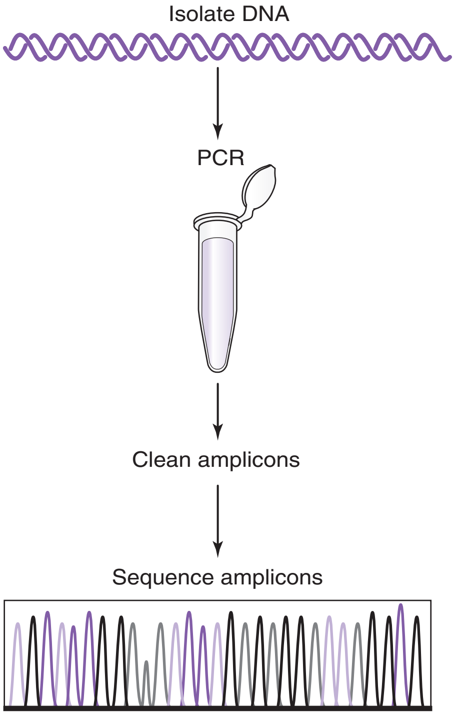
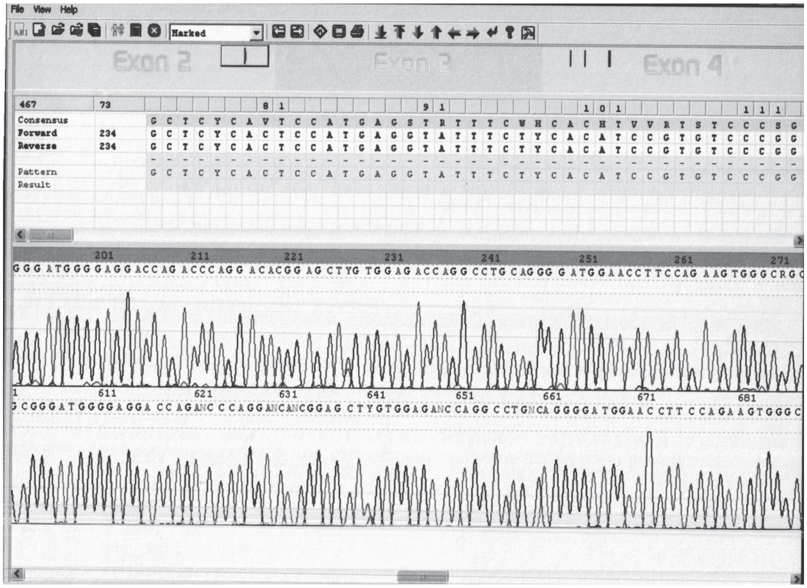


FIGURE 14.14 For Sanger sequence–based typing, HLA regions from patient DNA are amplified using the outer primers. The PCR products are then purified from unused PCR reaction components by alcohol precipitation or column or gel purification methods. The amplicons are then sequenced using the inner sequencing primers to detect polymorphisms.

of these problems. As with any sequencing assay, it is useful to sequence both strands of the test DNA.

Because HLA haplotypes are almost always heterozygous, Sanger sequencing results often yield a heterozygous pattern at the site of the polymorphism (Fig. 14.15). Phasing or placement of alleles on the same (maternal or paternal) chromosome in diploid organisms is difficult in Sanger typing. Also, ambiguous allele assignments revealed in rare alleles found by high-resolution sequencing require additional testing and delay results. Ambiguities are produced by a lack of sequencing of all polymorphic positions and/or lack of phasing. Therefore, exact high-resolution HLA typing is increasingly difficult with the growing number of HLA alleles not easily distinguished from one another. Commercial systems including PCR and sequencing reaction mixes and software programs for interpretation of results have been designed to address these issues.

Next-generation sequencing (NGS) offers the advantages of expanding the regions to be sequenced in phase while lowering ambiguities. NGS HLA panels cover the entire HLA-A, HLA-B, and HLA-C genes for class I and the entire HLA-DQB1, HLA-DQA1, and HLA-DPA1 genes and exon 2 of the HLA-DPB1 for class II. These regions are covered by long-range PCR, which allows accurate phasing of heterozygous alleles. Two technologies account for most HLA sequencing by



```
467        73          8 1               9 1                 1 0 1                 1 1 1
Consensus      G C T C Y C A V T C C A T G A G S T R T T T C W H C A C H T V V R T S T C C C S G
Forward    234 G C T C Y C A C T C C A T G A G G T A T T T C T Y C A C A T C C G T G T C C C G G
Reverse    234 G C T C Y C A C T C C A T G A G G T A T T T C T Y C A C A T C C G T G T C C C G G
               - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Pattern        G C T C Y C A C T C C A T G A G G T A T T T C T Y C A C A T C C G T G T C C C G G
Result

       201        211       221        231        241        251        261        271
GGGATGGGGAGGACCAGACCCAGGACACGGAGCTYGTGGAGACCAGGCCTGCAGGGGATGGAACCTTCCAGAAGTGGGCRG

1      611        621       631        641        651        661        671        681
GCGGGATGGGGAGGACCAGANCCCAGGANCANCGGAGCTYGTGGAGANCCAGGCCTGNCAGGGGATGGAACCTTCCAGAAGTGGGC
```

FIGURE 14.15 For a sequence-based typing result for the HLA-A gene, software is designed to analyze the sequence of patient DNA, compare it with the consensus sequence, and identify the specific alleles by sequence polymorphisms. *(Sequence courtesy of Christin Braun, Rush University Medical Center.)*

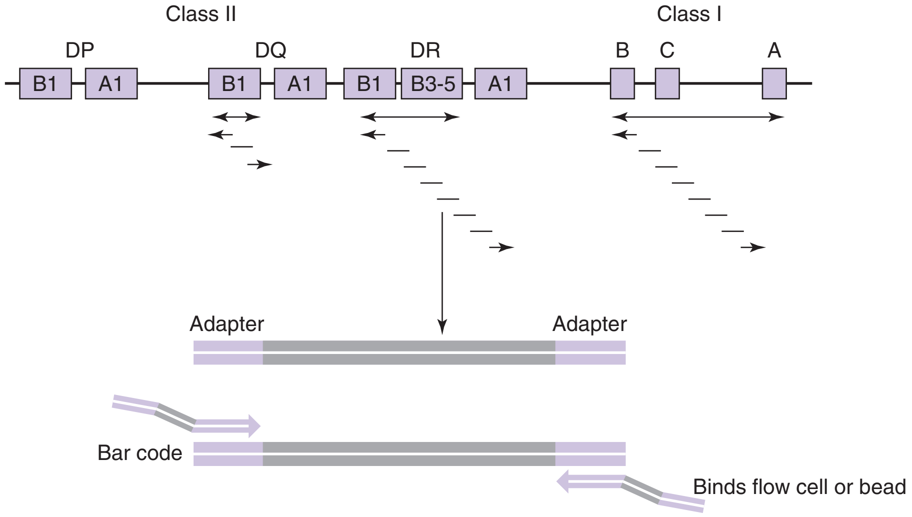


FIGURE 14.16 Next-generation sequencing libraries are produced from MHC regions selected by long-range PCR (double arrows). These 7- to 10-kb fragments are fragmented or amplified for generation of shorter fragments. The fragments are ligated to adapters and then tagged by PCR using primers tailed with bar codes for patient identification and sequencing primer-binding sites. In dye terminator sequencing, sequencing reactions are performed from polonies immobilized on a sequencing chip (flow cell). For ion conductance, sequencing reactions are performed on beads in wells of the sequencing chip.

NGS, reversible dye terminator sequencing[28] and ion conductance.[29]

For NGS typing, HLA genes are first amplified by long-range PCR (Fig. 14.16). The resulting PCR products range from 4 to 10 kb in size. The sequencing library is prepared from the amplified fragments by enzymatic fragmentation to products less than 1 kb in size and addition of secondary primer-binding sites. These sites are used to amplify the fragments with primers carrying patient-specific sequences (bar codes) and recognition sites for immobilized sequencing primers (Fig. 14.16). The bar codes allow sequencing of multiple patient samples simultaneously, with their sequences assigned to each patient by the bases in the bar code. Depending on the type of sequencer, the sequencing primers may be attached to beads or directly to a chip or flow cell.

After amplification, sequencing proceeds directly on the flow cell. Sequencing by ion conductance is performed in wells on the chip, each accommodating a bead on which each library fragment has been amplified. Library concentration and sequence quality are monitored in the course of sequencing. As with Sanger-based NGS, instrument software can analyze the sequence data. Specialized HLA analysis software is then used to assign haplotypes by comparison of the alleles to consensus sequences in the IMGT/HLA database of alleles.

Other DNA-Based Methods

Almost any method that can determine DNA sequence or detect specific sequences or sequence differences can be applied to DNA-based HLA typing. There are several variations on SSOP, SSP, and SBT, such as the nested PCR-SSP[30] proposed for high-sensitivity HLA typing. Another example is allele-specific nested PCR-SSP, which has been applied to subtyping of highly polymorphic alleles.

Heteroduplex (HD) analysis was used for assessment of compatible bone marrow donors at the HLA-DR and HLA-DP loci.[31] Reference strand conformation polymorphism was a variation on the standard HD analysis; sample amplicons were mixed with fluorescently labeled reference DNA of known allele sequence before denaturation and renaturation to form HD. The homoduplexes and HD formed between the specimen amplicons and the reference strand were then resolved by capillary gel electrophoresis.

Single-strand conformation polymorphism (SSCP) was also applied to HLA-A, DR, DQ, and DP typing and subtyping, sometimes coupled with allele-specific PCR. High-performance liquid chromatography (HPLC) was proposed for HLA typing as well. Conformation analyses, such as HD, SSCP, and HPLC, however, are limited by the complexity of the raw data and the strict demands on reactions and electrophoresis conditions; they are not routinely used in the clinical laboratory.

Other DNA-based typing approaches include array technology[32] and pyrosequencing.[33] No one method is without disadvantages with respect to technical demands, cost, or time consumption. To date, SSP, SSOP, Sanger sequencing, and NGS are the DNA-based methods used in most clinical laboratories.

TABLE 14.6 Resolution of HLA Typing Methods

Low Resolution	Intermediate Resolution	High Resolution
CDC (serology)	PCR-SSP	PCR-SSP
PCR-SSP	PCR-SSOP	PCR-SSOP
PCR-SSOP	PCR-RFLP	SSP-PCR + PCR-RFLP
		SSOP-PCR + SSP-PCR
		SBT

Combining Typing Results

At the DNA level, HLA polymorphisms differ from one another by as little as a single nucleotide base. Serological typing does not always distinguish subtle genetic differences between types, and also requires the proper specimen. A specimen consisting of mostly T cells, for example, from a patient treated with chemotherapy will not provide B cells (which carry class II antigens) for testing of class II haplotypes. In contrast, DNA-based typing is not limited by specimen type because all cells have the same HLA haplotype at the DNA level, regardless of whether the cell type expresses the antigens. Furthermore, synonymous DNA changes and polymorphisms outside of the protein-coding regions may not alter antigenicity at the protein level.

For DNA-based methods, the design of molecular methods (primer and probe selection) determines their level of resolution. SSOP and SSP methods require specific primers and/or probes for each particular HLA type. Only those HLA types included in a given probe or primer set, therefore, will be identified. Sanger sequence–based typing will only identify alleles included in the amplified regions that are sequenced. NGS provides almost complete coverage of HLA loci with better accuracy; however, this technology has yet to be implemented in many laboratories.

Results from serological and DNA-based methods can be combined to improve resolution and further define HLA types. Sequential use of SSP-PCR and PCR-RFLP or SSOP-PCR and SSP-PCR increases the typing resolution of DNA-based tests. Serological testing can be used to clarify or confirm the phenotype of alleles detectable by DNA-based methods. The resolution of results from various methods, therefore, reflects a range of resolution levels (Table 14.6). The choice of method will depend on the demand for high- or low-resolution typing.

HLA Test Discrepancies

HLA typing may produce discrepant results, especially if different methods are used to assess the same specimen. The most common discrepancies are those between serology and molecular testing results.

DNA sequence changes do not always affect protein epitopes. A serology type may represent several alleles at the DNA level. Also, a serology type may look homozygous (match to only one antibody) where the DNA alleles are heterozygous, the second allele not being recognized by serology. For example, a serology type of A2 is determined to be A*02, A*74 at the DNA level.

Discrepancies also arise when HLA types assigned to parent alleles based on DNA sequence homology differ from serology results that detect the same allele as a new antigen. For example, a DNA allele of B*40:05 is detected as a new antigen by serology and named B50. Similarly, split alleles (subtypes of serologically defined antigens) can differ between DNA and serology typing due to the cross reactivity of antibodies used to define the HLA antigens.

The identification of new alleles can result in ambiguities and discrepant retyping results based on the recognition of new alleles that were not defined at the time of an initial typing. These discrepancies may be difficult to resolve, especially if the original typing data are not accessible. Sequence ambiguities can be resolved by NGS, if available to the laboratory.

Coordination of HLA Test Methods

The choice of the appropriate method and resolution of HLA testing is influenced by the type of transplant. For solid organ transplants, antibody screening and crossmatching of recipient serum against donor antigens are routinely performed, although not always before transplantation. Pretransplant HLA typing is often determined for kidney and pancreas transplants because the extent of HLA matching is directly proportional to the time of survival of the donor organ. Given the circumstances under which heart, lung, and liver transplants are performed, testing is frequently performed after the transplant. HLA typing for solid organs is usually at the low-resolution serology level, although for heart and lung transplants, as with kidney transplants, studies have shown that matching HLA types are beneficial for organ survival.[34] For stem cell and bone marrow transplants, typing to high resolution (specific alleles) is preferred in order to decrease the risk of rejection and to avoid GVHD.[35] NGS has provided typing up to 4-digit allelic levels for this purpose.[36]

ADDITIONAL RECOGNITION FACTORS

Minor Histocompatibility Antigens

Any donor protein that can be recognized as nonself by the recipient immune system can potentially affect engraftment. Proteins outside the MHC that influence graft failure are called **minor histocompatibility antigens (mHags).** These antigens are the suspected cause of GVHD and graft rejection in MHC-identical transplants.[37,38] Conversely, mHags may promote graft-versus-tumor (GVT) effects as well.[38] Identification of tissue-specific mHags may allow inhibition of GVHD while increasing GVT activity.

The H-Y antigen was the first characterized mHag.[39] Molecular methods have led to the characterization of additional mHags.[40] Evaluation of mHags in stem cell transplants can be carried out by molecular methods such as SSP-PCR, sequencing, or even genome-wide correlations of mHag sequences with single-nucleotide polymorphisms from the HapMap project.[41,42]

Nonconventional MHC Antigens

Located within the MHC locus are the MHC class I–related *MICA* and *MICB* genes. Three pseudogene fragments, *MICC, MICD,* and *MICE,* are also found within the class I region. The products of the *MICA* and *MICB* genes, along with those of the retinoic acid early transcript (RAET) gene cluster, located on the long arm of chromosome 6 (6p21.3), bind to the receptor NKG2D on natural killer (NK) cells (killer cell lectin-like receptor, subfamily K, number 1 or KLRK1). A soluble isoform of MICA (sMICA) has the potential to suppress NKG2D-mediated host innate immunity by promoting degradation of the receptor-ligand complex. These gene products participate in immune reactions against abnormal cells such as tumor cells through control of NK cells and cytotoxic T lymphocytes (CTLs) expressing the γδ T-cell receptor. Virus- or bacteria-infected cells may also be recognized and eliminated in part by this system.

The *MICA* and *MICB* genes are highly polymorphic. Approximately 107 *MICA* and 42 *MICB* alleles have been reported. In contrast to the MHC class I alleles, polymorphisms in the *MIC* genes are distributed throughout the coding regions, with no hypervariable regions. Anti-MIC antibodies have been detected after organ transplantation, similar to anti-HLA alloantibodies, supporting a role for these gene products in organ rejection.[43]

Killer Cell Immunoglobulin-Like Receptors

NK cells and some memory T cells express **killer cell immunoglobulin-like (KIR)** proteins. The effect of these proteins was first observed as **"hybrid resistance"** in mice.[44] In these experiments, mice with compromised immune systems were still capable of rejecting grafts from unrelated mice. That is, graft rejection still occurred, even in the absence of a functional immune system. The KIR proteins have been proposed as one source of nonself recognition outside of the MHC. The KIR proteins interact with HLA antigens, specifically recognizing HLA-A, HLA-B, and HLA-C (class I)

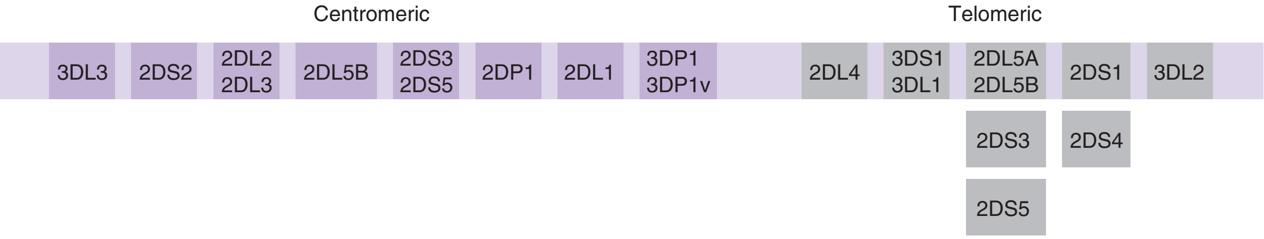


FIGURE 14.17 The KIR gene cluster includes a centromeric and a telomeric fragment. Gene content varies from one person to another. A KIR haplotype can contain from 8 to 14 genes in different combinations and gene orders. For example, a haplotype may have a 2DL2 or 2DL3 gene between the 2DS2 and 2DL5B genes, or the order of 2DS1 and 2DS4 can be 2DS1–2DS4 or 2DS4–2DS1.

molecules. KIR proteins are also expressed on myelomonocytic lineage cells (leukocyte immunoglobulin-like receptor) and other leukocytes (leukocyte-associated immunoglobulin-like receptor). A cluster of genes coding for these receptor proteins has been found on chromosome 19q13.4, the **leukocyte receptor cluster** (Fig. 14.17).

Recipient KIR may participate in graft rejection and donor KIR in GVHD.[45] Treatment for GVHD with cyclophosphamide after transplant may also inhibit the GVT effect of NK cells.[46] Specific interactions between KIR and HLA genes are listed in Table 14.7. Just as with mHags, assessment of polymorphisms in KIR may be added to donor selection criteria in stem cell and bone marrow transplants, especially with unrelated donors. In contrast to HLA typing, testing for KIR is aimed at finding donors and recipients who do *not* match. KIR alleles are typed by SSO using bead array technology and by NGS.[47]

TABLE 14.7 KIR and HLA Gene Interactions

KIR	HLA Specificities
2DL1, 2DS1 (2DS4)	Cw*02, *03:10, *03:04, *03:05, *03:06, *03:07, *03:15, *07:07, *07:09, *12:04, *12:05, *15 (except *15:07), *16:02, *17, *18
2DL2, 2DL3 (2DS2)	Cw*01, *03 (except *03:07), *03:10, *03:15, *07 (except *07:07, *07:09), *07:08, *12 (except *12:04, *12:05), *13, *14, *15:07, *16 (except *16:02)
2DL5	Unknown
3DL1	Bw4
3DL2	A3, A11
3DL3	Unknown
2DS3	Unknown
2DS5	Unknown
3DS1	Bw4

MHC DISEASE ASSOCIATION

Genetic diseases caused by single-gene disorders obey Mendelian laws. Their phenotypes are either **dominant** or **recessive** and are inherited in a predictable manner, as illustrated in pedigrees. Most diseases, however, are not caused by a single genetic lesion and therefore have complex segregation patterns. Multiple genes, epigenetics, and environmental factors combine to bring about these disease states. For diseases such as diabetes, high blood pressure, and certain cancers, genetic analysis yields results in terms of predisposition, probability, and risk of disease.

Autoimmune diseases, which affect 4% of the population, fall in this category. At least one of the genetic factors involved in autoimmunity is linked to the MHC because autoimmune diseases have MHC associations.[48,49] Rheumatoid arthritis, multiple sclerosis, diabetes mellitus type 1, and systemic lupus erythematosus are associated with particular HLA haplotypes. Determination of a disease-associated HLA haplotype aids in diagnosis or prediction of disease predisposition.

The presence of a haplotype associated with disease is not diagnostic on its own, however. An example is the HLA-B27 type found in all cases of ankylosing spondylitis. HLA-B27 is also the most frequently found HLA-B allele. Many people, therefore, with the B27 allele do not have this condition. Other associations, such as narcolepsy being strongly associated with HLA-DQB1*06:02, may arise from involvement of presentation of an autoantigen to CD4+ T cells in the context of specific HLA types and subsequent autoimmunity.[50] The laboratory is usually asked to perform HLA typing in persons showing disease symptoms or family histories of such disease to aid or confirm the diagnostic decisions.

Normal states are also controlled by multiple genetic and environmental factors. The genes for olfactory sensation, histone genes, genes encoding transcription factors, and the butyrophilin (a constituent of milk in mammals) gene cluster are found in the MHC. The MHC may play a role not only in the predisposition to disease but also in protection from disease, including some infectious diseases, notably HIV.[51] In this role, certain HLA types are considered "protective." It has been suggested that HLA type be considered in antiviral therapy trials.[52]

SUMMARY OF LABORATORY TESTING

Standard HLA testing includes a range of methods and test designs (Table 14.8). Determining donor and recipient compatibility is the primary goal of pretransplant testing, especially for bone marrow transplant from unrelated donors. The extent of HLA matching will increase the prospect of successful transplantation with minimal GVHD. For example, the NMDP requests HLA typing results for HLA-A, HLA-B, and HLA-C class I antigens and DRB1, DQB1, and DPB1 class II antigens for prospective bone marrow donors. Advances in SBT using NGS have increased the resolution of typing and limited the number of ambiguities that formerly had to be resolved by other methods, thus shortening the reporting time for typing.

TABLE 14.8 Summary of Testing for Pre- and Post-Transplant Evaluation

Purpose	Test Design	Methods
Determine HLA type	Determine HLA types with standard references or by DNA sequence	Serology
		DNA-based typing
Determine serum antibody status	Serum screening against known HLA antibodies	CDC
		ELISA
		Flow cytometry
Crossmatching	Compare serum of recipient with that of prospective donors	CDC
		ELISA
		Flow cytometry
Determine T-cell-mediated cytotoxicity between donor and recipient	Alloreactive T-cell characterization and quantitation	MLC
		Cytokine production

Advanced Concepts

HLA types are also associated with the absence of inherited disease. Protection from genetic diseases is naturally enhanced by "hybrid vigor" or avoidance of inbreeding. The MHC may bring about such natural protection by control of mating selection to ensure genetic mixing. For example, studies with mice have shown that olfactory sensation (sense of smell) may play a role in mate selection. In these studies, female mice were able to distinguish male mice with H-2 alleles different from their own.[53] This supports the idea of selective pressure to mate with HLA-dissimilar partners to increase genetic diversity. Another study in humans showed that women found the scent of T-shirts worn by some male subjects pleasant and others unpleasant. The odors detected as pleasant were from HLA-dissimilar males. The women were apparently able to distinguish HLA types different from their own through the sweat odors of male test subjects.[54]

Combinations of molecular tests, serum screening, and crossmatching further define acceptable HLA mismatches, and the results are used to prevent hyperacute (almost immediate) rejection in organ transplants. Identification of HLA types associated with disease also provides important information, especially for disease caused by multiple genetic and environmental factors. Finally, MLC, although not generally part of routine clinical testing, can predict cellular factors involved in rejection.

Laboratory results are key for the selection of compatible donors, post-transplant evaluation, selection of optimal treatment strategy, and evaluation of genetic disease predisposition. Molecular analysis has significantly increased the ease and ability to detect subtle differences in the MHC and associated regions. Successful organ transplantation will increase as databases grow and instrumentation, automation, and bioinformatics advance. Nucleic acid analysis has contributed greatly to more definitive analysis in this area of laboratory testing.

Case Study 14.1

A 50-year-old man complained of digestive disorders and what he presumed was an allergy to several foods. He consulted his physician, who collected blood samples for laboratory testing. The man was in a high-risk group for certain diseases, notably celiac disease, which would produce symptoms similar to those experienced by this patient. A specimen was sent to the laboratory to test HLA-DQA and HLA-DQB alleles. Almost all (95%) people with celiac disease have the DQA1*05:01 and DQB1*02:01 alleles, compared with 20% of the general population. A second predisposing heterodimer, DQ8, is encoded by the DQA1*03:01 and the DQB1*03:02 alleles. Most of the 5% of celiac patients who are negative for the DQ2 alleles display the DQ8 alleles. Serological results to detect the predisposing antigens, however, were equivocal. An SBT test was performed. The indicated alleles were detected by sequence analysis. The sequence results showed the following alleles at DQA1 and DQB1 loci:

$$\text{DQA1*05:01} \qquad \text{DQB1*02:01, DQB1*04:01}$$

QUESTION: *Is it possible that this man has celiac disease based on his HLA haplotype?*

Case Study 14.2

A 43-year-old man consulted his physician about a lump on his neck and frequent night sweats. A biopsy of the mass in his neck was sent to the pathology department for analysis. An abnormally large population of CD20-positive lymphocytes was observed by morphological examination. Flow-cytometry tests detected a monoclonal B-cell population with coexpression of CD10/CD19 and CD5. This population was 88% kappa and 7% lambda. The results were confirmed by the observation of a monoclonal immunoglobulin heavy-chain gene rearrangement that was also monoclonal for kappa light-chain gene rearrangement. The patient was initially treated with standard chemotherapy, but the tumor returned before the therapeutic program was completed. The tumor persisted through a second treatment with stronger chemotherapy plus local irradiation. Nonmyeloablative bone marrow transplant was prescribed. To find a compatible donor, the man's HLA type and the types of five potential donors were compared. HLA-A and HLA-B types were assessed by serology, and HLA-DR type was determined by SSP-PCR and SSOP. The typing results are shown in the accompanying table.

Recipient	**Donor 1**	**Donor 2**	**Donor 3**	**Donor 4**	**Donor 5**
A*03:02	A*66:01/04	A*26:01	A*03:02	A*31;03/04	A*26;10
A*26:10	A*03:02	A*43:01		A*03;09	A*0309
B*39:06	B*07:11	B*15:28	B*38:01	B*27;02	B*35;08
B*53:07	B*57:01	B*39:19	B*39:01	B*3501	
DRB1*08:01	DRB1*07:01	DRB1*13:17	DRB1*08:09	DRB1*07;01	DRB1*13;17
	DRB1*13:17	DRB1*04:22		DRB1*11:17	DRB1*07;01

QUESTIONS:

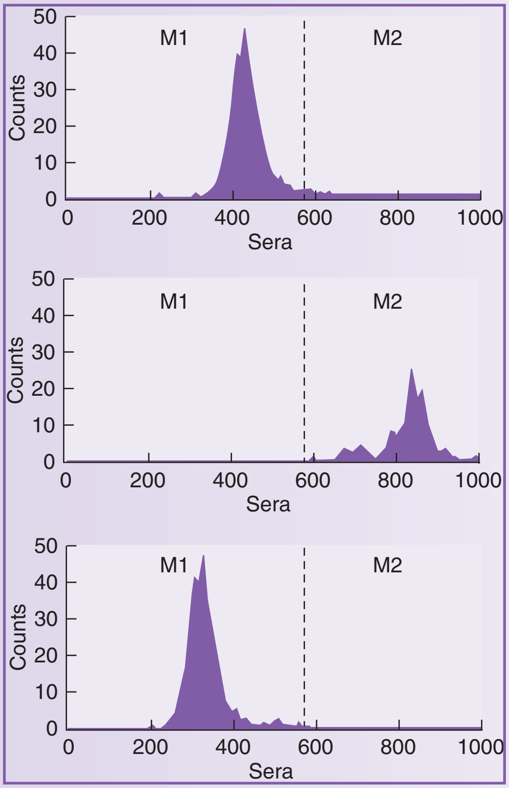
1. *Which of the five donors is the best match for this patient? Which mismatches are acceptable?*
2. *Using the donor bone marrow that matches most closely, is there more of a chance of graft rejection or GVHD?*

Case Study 14.3

A 19-year-old woman reported to a local clinic with painful swelling in her face. Routine tests revealed dangerously high blood pressure that warranted hospitalization. Further tests were performed, which led to a diagnosis of systemic lupus erythematosus. Due to complications of this disease, her kidney function was compromised, and she would eventually suffer kidney failure. With an alternative of lifelong dialysis, a kidney transplant was recommended. Class I PRAs were assessed at 0% PRA for class I. An additional screen for class II PRA was performed by flow cytometry.

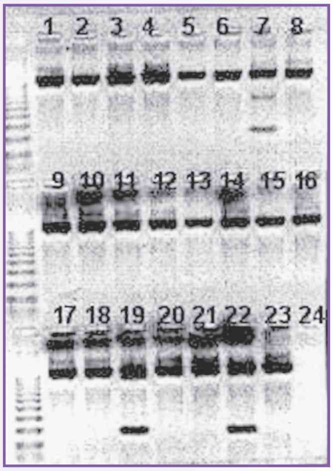
Continued on following page

Case Study 14.3 *(Continued)*



Flow-cytometry analysis of HLA class II antigens in recipient serum. Top = negative control; middle = positive control; bottom = patient serum.

Class I and class II HLA typing was performed by SSP-PCR. An autocytotoxic crossmatch was also performed, revealing T-cell- and B-cell-positive antibodies, probably related to the lupus. The young woman's mother volunteered to donate a kidney to her daughter. Mother and daughter had matching blood group antigens and HLA-DR antigens, which are the most critical for a successful organ transplant. The results from the typing tray are shown in the following figure:



SSP-PCR results from an HLA-A typing tray for the daughter. Alleles were identified in lanes 7, 19, and 22.

SSP-PCR results for HLA-A, HLA-B, and HLA-DR and their serological equivalents are shown in the accompanying table.

Crossmatching of the daughter's serum and the mother's cells was performed by cytotoxicity and flow cytometry. Both test results were negative.

QUESTIONS:

1. *Is the mother a good match for the daughter?*
2. *Based on the antibody and crossmatching studies, what is the risk of rejection? Before performing the HLA studies, how many of the daughter's antigens would be expected to match those of her mother?*

Daughter		**Mother**	
SSP-PCR	Serology	SSP-PCR	Serology
A*24:19, A*3401	A24(9), 34(10)	A*11:04, A*24:19	A11, A24(9)
B*08:02, B*56:03	B8, B22	B*08:04, B*56:03	B8, B22
DRB1*04:22, DRB1*14:11	DR4, DR14(6)	DRB1*04:22, DRB1*13:17	DR4, DR13(6)

STUDY QUESTIONS

1. Which of the following is a high-resolution HLA typing result?
   - a. B27
   - b. A*02:02–02:09
   - c. A*02:12
   - d. A*26:01/A*26:05/A*26:01/A*26:15

2. Which of the following is a likely haplotype from parents with A25,Cw10,B27/A23,Cw5,B27 and A17,Cw4,B10/A9,Cw7,B12 haplotypes?
   - a. A25,Cw10,B27
   - b. A25,Cw5,B27
   - c. A23,Cw4,B12
   - d. A17,Cw4,B27

3. Upon microscopic examination, over 90% of cells are translucent after a CDC assay. How are these results scored according to the ASHI rules?

4. An HLA-A allele is a CTC to CTT (leu → leu) change at the DNA level. How is this allele written?
   - a. HLA-A*02
   - b. HLA-A*02:01
   - c. HLA-A2
   - d. HLA-A*02N

5. A candidate for kidney transplant has a PRA of 75%. How will this affect eligibility for immediate transplant?

6. An SSOP probe recognizes HLA-DRB*03:01–03:04. Another probe recognizes HLA-DRB*03:01/03:04, and a third probe hybridizes to HLA-DRB*03:01–03:03. Test specimen DNA hybridizes to all except the third probe in a reverse dot-blot format. What is the HLA-DRB type of the specimen?

7. What is the relationship between alleles HLA-A*10 and HLA-A*26(10)?

8. A CDC assay yields an 8 score for sera with the following specificities: A2, A28 and A2, A28, B7, and a 1 score for serum with an A2 specificity. What is the HLA-A type?

9. HLA-DRB1*15:01 differs from DRB1*01:01 by a G to C base change. If the sequence surrounding the base change is …GGGTGCGGTTGCTGGAAAGAT… (DRB1*01:01) or …GGGTGCGGTTCCTGGAAAGAT… (DRB1*15:01), which of the following would be the 3′ end of a sequence-specific primer for detection of DRB1*15:01?
   - a. …ATCTTTCCAGGAACCC
   - b. …ATCTTTCCAGCAACCC
   - c. …ATCTTTCCAGC
   - d. …ATCTTTCCAGG

10. The results of an SSP-PCR reaction are the following: lane 1, one band; lane 2, two bands; lane 3, no bands. If the test includes an amplification control multiplexed with the allele-specific primers, what is the interpretation for each lane?

References

1. Bain B, Vaz MR, Lowenstein L. The development of large immature mononuclear cells in mixed lymphocyte cultures. *Blood* 1964;23:108–116.
2. Bach F, Hirschhorn K. Lymphocyte interaction: a potential histocompatibility test in vitro. *Science* 1964;142:813–814.
3. Jones E, Goodfellow PN, Bodmer JG, Bodmer WF. Serological identification of HL-A linked human “Ia-type” antigens. *Nature* 1975;256:650–652.
4. Park M, Terasaki PI, Bernoco D, Iwaki Y. Evidence for a second B-cell locus separate from the DR locus. *Transplantation Proceedings* 1978;10:823–828.
5. Bover K. Homoisotransplantation van Epidermis bei eineigen Zwillingen. *Beitrage Zur Klinischen Chirurgica* 1927;141:442–447.
6. Gorer P. The genetic and antigenic basis for tumor transplantation. *Journal of Pathology and Bacteriology* 1937;44:691–697.
7. Snell G. Methods for the study of histocompatibility genes. *Journal of Genetics* 1948;49:87–108.
8. Inoko H, Trowsdale J. Linkage of TNF genes to the HLA-B locus. *Nucleic Acids Research* 1987;15:8957–8962.
9. Ruddy DA, Kronmal GS, Lee VK, Mintier GA, Quintana L, Domingo R, Meyer NC, Irrinki A, McClelland EE, Fullan A,

Mapa FA, Moore T, Thomas W, Loeb DB, Harmon C, Tsuchihashi Z, Wolff RK, Schatzman RC, Feder JN. A 1.1-Mb transcript map of the hereditary hemochromatosis locus. *Genome Research* 1997;7:441–456.

10. Feder J, Gnirke A, Thomas W, Tsuchihashi Z, Ruddy DA, Basava A, Dormishian F, Domingo R, Ellis MC, Fullan A, Hinton LM, Jones NL, Kimmel BE, Kronmal GS, Lauer P, Lee VK, Loeb DB, Mapa FA, McClelland E, Meyer NC, Mintier GA, Moeller N, Moore T, Morikang E, Prass CE, Quintana L, Starnes SM, Schatzman RC, Brunke KJ, Drayna DT, Risch NJ, Bacon BR, Wolff RK. A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. *Nature Genetics* 1996;13:399–408.
11. Blees A, Januliene D, Hofmann T, Koller N, Schmidt C, Trowitzsch S, Moeller A, Tampé R. Structure of the human MHC-I peptide-loading complex. *Nature* 2017;551:525–528.
12. Barton J, Barton JC, Acton RT. Longer survival associated with HLA-A*03, B*14 among 212 hemochromatosis probands with HFE C282Y homozygosity and HLA-A and -B typing and haplotyping. *European Journal of Haematology* 2010;85:439–447.
13. Gao X, Bashirova A, Iversen AK, Phair J, Goedert JJ, Buchbinder S, Hoots K, Vlahov D, Altfeld M, O'Brien SJ, Carrington M. AIDS restriction HLA allotypes target distinct intervals of HIV-1 pathogenesis. *Nature Medicine* 2005;11:1290–1292.
14. de Silva T, Leligdowicz A, Carlson J, Garcia-Knight M, Onyango C, Miller N, Yindom LM, Hué S, Jaye A, Dong T, Cotten M, Rowland-Jones SL. HLA-associated polymorphisms in the HIV-2 capsid highlight key differences between HIV-1 and HIV-2 immune adaptation. *AIDS* 2018;32(6):709–714,
15. Dunn P, Day S, Williams S, Bendukidze N. HLA-DQB1 sequencing-based typing using newly identified conserved nucleotide sequences in introns 1 and 2. *Tissue Antigens* 2005;66:99–106.
16. Owerbach D, Lernmark A, Rask L, Peterson PA, Platz P, Svejgaard A. Detection of HLA-D/DR-related DNA polymorphism in HLA-D homozygous typing cells. *Proceedings of the National Academy of Sciences* 1983;80:3758–3761.
17. Marsh S, Bodmer JG, Albert ED, Bodmer WF, Bontrop RE, Dupont B, Erlich HA, Hansen JA, Mach B, Mayr WR, Parham P, Petersdorf EW, Sasazuki T, Schreuder GM, Strominger JL, Svejgaard A, Terasaki PI. Nomenclature for factors of the HLA system, 2000. *Tissue Antigens* 2001;57:263–283.
18. Marsh S, Albert ED, Bodmer WF, Bontrop RE, Dupont B, Erlich HA, Geraghty DE, Hansen JA, Hurley CK, Mach B, Mayr WR, Parham P, Petersdorf EW, Sasazuki T, Schreuder GMT, Strominger JL, Svejgaard A, Terasaki PI, Trowsdale J. Nomenclature for factors of the HLA system, 2004. *Human Immunology* 2005;66:571–636.
19. Holdsworth R, Hurley CK, Marsh SGE, Lau M, Noreen HJ, Kempenich JH, Setterholm M, Maiers M. The HLA Dictionary 2008: a summary of HLA-A, -B, -C, -DRB1/3/4/5, -DQB1 alleles and their association with serologically defined HLA-A, -B, -C, -DR and -DQ antigens. *Tissue Antigens* 2009;73:95–170.
20. Peña J, Fitzpatrick D, Saidman SL. Complement-dependent cytotoxicity crossmatch. *Methods in Molecular Biology* 2013; 1034:257–283.
21. Irure J, Asensio E, Rodrigo E, Romón Í, Gómez J, Arias M, López-Hoyos M, San Segundo D. Improvement in the definition of anti-HLA antibody profile in highly sensitized patients. *PLoS One* 2017;12:e0171463.
22. Kousoulas L, Vondran FW, Syryca P, Klempnauer J, Schrem H, Lehner F. Risk-adjusted analysis of relevant outcome drivers for patients after more than two kidney transplants. *Journal of Transplantation* 2015;712049:1–9.
23. Chu S, Chien KH, Lin HH, Wu WH, Jian JY, Tzeng WF, Chiou TJ. Using marker gene analysis instead of mixed lymphocyte reaction assay for identification of functional CD4+FOXP3+ regulatory T cells. *Biotechnology Letters* 2018;40(3):535–542.
24. Faridi P, Purcell AW, Croft NP. In immunopeptidomics we need a sniper instead of a shotgun. *Proteomics* 2018;18(12):e1700464.
25. Creech A, Ting YS, Goulding SP, Sauld JF, Barthelme D, Rooney MS, Addona TA, Abelin JG. The role of mass spectrometry and proteogenomics in the advancement of HLA epitope prediction. *Proteomics* 2018;18(12):e1700259.
26. Flomenberg N, Baxter-Lowe LA, Confer D, Fernandez-Vina M, Filipovich A, Horowitz M, Hurley C, Kollman C, Anasetti C, Noreen H, Begovich A, Hildebrand W, Petersdorf E, Schmeckpeper B, Setterholm M, Trachtenberg E, Williams T, Yunis E, Weisdorf D. Impact of HLA class I and class II high-resolution matching on outcomes of unrelated donor bone marrow transplantation: HLA-C mismatching is associated with a strong adverse effect on transplantation outcome. *Blood* 2004;104:1923.
27. Cao K, Chopek M, Fernandez-Vina MA. High and intermediate resolution DNA typing systems for class I *HLA-A, B, C* genes by hybridization with sequence-specific oligonucleotide probes (SSOP). *Reviews in Immunogenetics* 1999;1:177–208.
28. Gandhi M, Ferriola D, Lind C, Duke JL, Huynh A, Papazoglou A, Mackiewicz K, Christiansen M, Dong W, Hsu S, Thomas D, Schneider B, Pierce E, Kearns J, Kamoun M, Monos D, Askar M. Assessing a single targeted next generation sequencing for human leukocyte antigen typing protocol for interoperability, as performed by users with variable experience. *Human Immunology* 2017;78:642–648.
29. Barone J, Saito K, Beutner K, Campo M, Dong W, Goswami CP, Johnson ES, Wang ZX, Hsu S. HLA-genotyping of clinical specimens using Ion Torrent-based NGS. *Human Immunology* 2015;76:903–909.
30. Song E, Chung HY, Joo SY, Roh EY, Seong MW, Shin Y, Park MH. Detection of HLA-DRB1 microchimerism using nested polymerase chain reaction and single-strand conformation polymorphism analysis. *Human Immunology* 2012;73:291–297.
31. Zhou L, Vandersteen J, Wang L, Fuller T, Taylor M, Palais B, Wittwer CT. High-resolution DNA melting curve analysis to establish HLA genotypic identity. *Tissue Antigens* 2004;64:156–164.
32. Bang-Ce Y, Xiaohe C, Ye F, Songyang L, Bincheng Y, Peng Z. Simultaneous genotyping of DRB1/3/4/5 loci by oligonucleotide microarray. *Journal of Molecular Diagnostics* 2005;7:592–599.
33. Entz P, Toliat MR, Hampe J, Valentonyte R, Jenisch S, Nernberg P, Nagy M. New strategies for efficient typing of HLA class-II loci DQB1 and DRB1 by using pyrosequencing. *Tissue Antigens* 2005;65:67–80.

34. Zachary A, Leffell MS. HLA mismatching strategies for solid organ transplantation—a balancing act. *Frontiers in Immunology* 2016;7:575–589.

35. Rocha V, Cornish J, Sievers EL, Filipovich A, Locatelli F, Peters C, Remberger M, Michel G, Arcese W, Dallorso S, Tiedemann K, Busca A, Chan KW, Kato S, Ortega J, Vowels M, Zander A, Souillet G, Oakill A, Woolfrey A, Pay AL, Green A, Garnier F, Ionescu I, Wernet P, Sirchia G, Rubinstein P, Chevret S, Gluckman E. Comparison of outcomes of unrelated bone marrow and umbilical cord blood transplants in children with acute leukemia. *Blood* 2001;97:2962–2971.

36. Zhou M, Gao D, Chai X, Liu J, Lan Z, Liu Q, Yang F, Guo Y, Fang J, Yang L, Du D, Chen L, Yang X, Zhang M, Zeng H, Lu J, Chen H, Zhang X, Wu S, Han Y, Tan J, Cheng Z, Huang C, Wang W. Application of high-throughput, high-resolution and cost-effective next generation sequencing-based large-scale HLA typing in donor registry. *Tissue Antigens* 2015;85:20–28.

37. Miklos D, Kim HT, Miller KH, Guo L, Zorn E, Lee SJ, Hochberg EP, Wu CJ, Alyea EP, Cutler C, Ho V, Soiffer RJ, Antin JH, Ritz J. Antibody responses to H-Y minor histocompatibility antigens correlate with chronic graft-versus-host disease and disease remission. *Blood* 2005;105:2973–2978.

38. van der Torren C, van Hensbergen Y, Luther S, Aghai Z, Rychnavská ZS, Slot M, Scherjon S, Kröger N, Ganser A, Weissinger EM, Goulmy E, Hambach L. Possible role of minor h antigens in the persistence of donor chimerism after stem cell transplantation; relevance for sustained leukemia remission. *PLoS One* 2015;10:1–18.

39. Goulmy E, Termijtelen A, Bradley BA, van Rood JJ. Alloimmunity to human H-Y. *Lancet* 1976;2:1206.

40. Oostvogels R, Lokhorst HM, Mutis T. Minor histocompatibility Ags: identification strategies, clinical results and translational perspectives. *Bone Marrow Transplant* 2016;51:163–171.

41. Wilke M, Pool J, den Haan JM, Goulmy E. Genomic identification of the minor histocompatibility antigen HA-1 locus by allele-specific PCR. *Tissue Antigens* 1998;52:312–317.

42. Spaapen R, de Kort RA, van den Oudenalder K, van Elk M, Bloem AC, Lokhorst HM, Mutis T. Rapid identification of clinical relevant minor histocompatibility antigens via genome-wide zygosity-genotype correlation analysis. *Clinical Cancer Research* 2009;15:7137–7143.

43. Bahram S, Inoko H, Shiina T, Radosavljevic M. MIC and other NKG2D ligands: from none to too many. *Current Opinion in Immunology* 2005;17:505–509.

44. Cudkowicz G, Bennett M. Peculiar immunobiology of bone marrow allografts. I. Graft rejection by irradiated responder mice. *Journal of Experimental Medicine* 1971;134:83–102.

45. Tran T, Mytilineos J, Scherer S, Laux G, Middleton D, Opelz G. Analysis of KIR ligand incompatibility in human renal transplantation. *Transplantation* 2005;80:1121–1123.

46. Russo A, Oliveira G, Berglund S, Greco R, Gambacorta V, Cieri N, Toffalori C, Zito L, Lorentino F, Piemontese S, Morelli M, Giglio F, Assanelli A, Stanghellini MTL, Bonini C, Peccatori J, Ciceri F, Luznik L, Vago L. NK cell recovery after haploidentical HSCT with posttransplant cyclophosphamide: dynamics and clinical implications. *Blood* 2018;131:247–262.

47. Maniangou B, Retière C, Gagne K. Next-generation sequencing technology a new tool for killer cell immunoglobulin-like receptor allele typing in hematopoietic stem cell transplantation. *Transfusion Clinique et Biologique* 2018;25:87–89.

48. Dendrou C, Petersen J, Rossjohn J, Fugger L. HLA variation and disease. *Nature Review Immunology* 2018;18(5):325–339.

49. Steck A, Bugawan TL, Valdes AM, Emery LM, Blair A, Norris JM, Redondo MJ, Babu SR, Erlich HA, Eisenbarth GS, Rewers MJ. Association of non-HLA genes with type 1 diabetes autoimmunity. *Diabetes* 2005;54:2482–2486.

50. Kornum B, Burgdorf KS, Holm A, Ullum H, Jennum P, Knudsen S. Absence of autoreactive CD4+ T-cells targeting HLA-DQA1*01:02/DQB1*06:02 restricted hypocretin/orexin epitopes in narcolepsy type 1 when detected by EliSpot. *Journal of Neuroimmunology* 2017;309:7–11.

51. Ndung'u T, Gaseitsiwe S, Sepako E, Doualla-Bell F, Peter T, Kim S, Thior I, Novitsky VA, Essex M. Major histocompatibility complex class II (HLA-DRB and -DQB) allele frequencies in Botswana: association with human immunodeficiency virus type 1 infection. *Clinical and Diagnostic Laboratory Immunology* 2005;12:1020–1028.

52. Park Y, Etemad B, Ahmed H, Naranbhai V, Aga E, Bosch RJ, Mellors JW, Kuritzkes DR, Para M, Gandhi RT, Carrington M, Li JZ. Impact of HLA class I alleles on timing of HIV rebound after antiretroviral treatment interruption. *Pathogens and Immunity* 2017;2:431–445.

53. Montag S, Frank M, Ulmer H, Wernet D, Gopel W, Rammensee HG. "Electronic nose" detects major histocompatibility complex-dependent prerenal and postrenal odor components. *Proceedings of the National Academy of Sciences* 2001;98:9249–9254.

54. Santos P, Schinemann JA, Gabardo J, Bicalho Mda G. New evidence that the MHC influences odor perception in humans: a study with 58 Southern Brazilian students. *Hormones and Behavior* 2005;47:384–388.

Chapter 15

Quality Assurance and Quality Control in the Molecular Laboratory

Outline

- **SPECIMEN HANDLING**
  - *Collection Tubes for Molecular Testing*
  - *Precautions*
  - *Holding and Storage Requirements*
- **TEST PERFORMANCE**
  - *Next-Generation Sequencing*
  - *Calibrators and Method Calibration*
  - *Controls*
- **QUALITY CONTROL**
- **QUALITY ASSURANCE**
- **INSTRUMENT MAINTENANCE**
  - *Instrument Calibration*
- **REAGENTS**
  - *Reagent Categories*
  - *Chemical Safety*
  - *Reagent Storage*
  - *Reagent Labeling*
- **PROFICIENCY TESTING**
- **DOCUMENTATION OF TEST RESULTS**
  - *Gene Nomenclature*
  - *Gene Sequencing Results*
- **REPORTING RESULTS**

Objectives

15.1 Describe proper specimen accession for molecular testing.
15.2 Describe the optimal conditions for holding and storage of specimens and nucleic acid.
15.3 Explain the basic components of molecular test performance, including quality assurance and controls.
15.4 Discuss instrument maintenance, repair, and calibration, particularly for instruments used in molecular analysis.
15.5 Describe recommendations for the preparation and use of reagents in the molecular laboratory.
15.6 Explain documentation and reporting of results, including gene sequencing results.

Congress passed the **Clinical Laboratory Improvement Amendments (CLIA)** in 1988 to establish quality testing standards to ensure consistent patient test results. CLIA specifies quality standards for proficiency testing, patient test management, quality control, personnel qualifications, and quality assurance for laboratories performing moderate- and/or high-complexity tests, including molecular testing. This chapter offers a brief overview of laboratory standards applied to molecular diagnostic tests.

SPECIMEN HANDLING

Molecular tests, like any clinical laboratory tests, require optimal specimen handling and processing for accurate and consistent test results. The success of a test procedure is affected by the age, type, and condition of specimens. Therefore, specimen collection, transport, and handling in the laboratory require careful attention.

Preanalytical variables, both controllable and uncontrollable, must be taken into account for proper interpretation of test results. **Preanalytical error** is the consequence of erroneous or misleading results caused by events that occur prior to sample analysis. To minimize preanalytical error and maximize control of preanalytical variables, the Clinical and Laboratory Standards Institute (CLSI, formerly known as the National Committee for Clinical Laboratory Standards) provides recommendations for the collection of specimens under standardized conditions.

Each laboratory will have requirements for specimen handling, but general policies apply to all specimen collection. A requisition or electronic test order should accompany the specimen. The condition of the specimen and, if necessary, the chain of custody is reviewed on receipt in the laboratory. If a specimen shows evidence of tampering or is hemolyzed, degraded, clotted, or otherwise compromised, the technologist must notify the supervisor. No specimen is accepted without proper labeling and identification on the specimen tube or container (placed by the person who collected the specimen), nor is a specimen accepted if the labeling on the specimen does not match that on the accompanying requisition. In addition to relevant patient identification, the test requisition includes the type of specimen material (e.g., blood or bone marrow), ordered test, date and time of collection, reason for testing, and the contact information (pager or telephone number) of the ordering physician. When required (for molecular genetics, forensics, or parentage testing), patient consent forms, ethnicity, photo identification of the individuals tested, patient label verification, and transfusion history or a pedigree may also be supplied with the test specimen. Forensic specimens may require a documented chain of custody. Bar coding of this information expedites specimen accession. The laboratory should have written procedures for documentation of specimen handling and accession. Accession books or electronic records are used to record the date of receipt, laboratory identifier, and pertinent patient information associated with the accession. If a specimen is unacceptable, the disposal or retention of the specimen is recorded in the patient report or laboratory quality assurance records. If the specimen cannot be tested, the ordering physician is notified.

If not processed immediately, specimens are maintained in secure areas with limited access under the appropriate conditions for the analyte being tested (Fig. 15.1). Care is taken to avoid contamination or mixing if specimens are aliquoted. Slides of thin sections used for in situ hybridization procedures are retained for

FIGURE 15.1 Biohazard stickers are required for cabinets, refrigerators, or freezers that contain potentially hazardous reagents or patient specimens. See Color Plate 10.

10 years or 20 years for neoplastic or constitutional analyses, respectively.

Molecular amplification methods have enabled laboratory professionals to perform nucleic acid–based testing on specimens with minimal cellular content, such as buccal cell suspensions and cerebrospinal fluids. These samples are centrifuged to collect the cells before DNA or RNA is extracted. For routine specimens tested by amplification methods, the entire specimen is often not used. In this case, or if more than one test is to be performed on the same specimen, cross-contamination of specimens is carefully avoided. This can most likely occur from pipetting carryover. Moreover, an aliquot removed from a specimen is never returned to the original tube or vessel. Molecular genetic tests may require dedicated specimens not shared with other tests.

Hemoglobin inhibits enzyme activity. Specimens received in the laboratory should therefore be inspected for visual signs of hemolysis. Hemoglobin and coagulants are removed effectively in most DNA and RNA isolation procedures; however, if white blood cell lysis has also occurred, DNA or RNA yield will be reduced. This could result in false-negative results in qualitative testing or inaccurate measurements in quantitative analyses. Buffers and resins have been designed to sequester anticoagulants or hemoglobin for more rapid nucleic acid isolation without the inhibitory effects of these substances.

Solid tissues are best analyzed from fresh or frozen samples (Fig. 15.2), especially for Southern blot or long-range polymerase chain reaction (PCR) methods that require relatively high-quality (long, intact) DNA. Surgical specimens designated for molecular studies, if not processed immediately, should be snap-frozen in liquid nitrogen. This process preserves both nucleic acid and most gene-expression patterns that may change upon tissue storage. Snap freezing is routinely performed in the surgical pathology laboratory because it is a common process for preserving tissue morphology for microscopic examination.

Fixed, paraffin-embedded tissues generally yield lower-quality DNA and RNA, depending on the type of fixative used, the amount of time of exposure of the tissue to the fixative, and how the specimen was handled prior to fixation. An advantage of fixed tissue is the use of adjacent slices of tissue that have been stained with hematoxylin and eosin (H&E) to identify particular tissue areas to be tested. For oncology testing, preanalytical review of the H&E slide by a pathologist is required to confirm adequacy and to designate what part of the tissue sample will be tested. Tumor tissue samples are not always well defined or may be infiltrated with lymphocytes. Therefore, the pathologist may also estimate the percentage of tumor cells in the area to be tested.

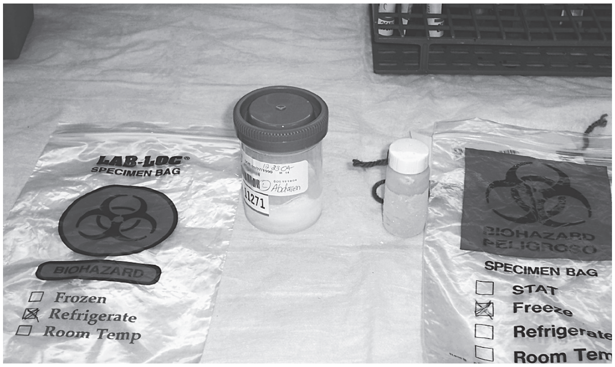


FIGURE 15.2 Tissue is received in the laboratory in fresh *(left)* or frozen *(right)* form. Fresh tissue may be supplied as a specimen on gauze or another substrate or in saline. The small vial shown on the right contains tissue that was flash-frozen in isopentane. The vial is held immersed in liquid nitrogen, in nitrogen vapors, or in a –70°C freezer.

PCR and reverse transcription PCR (RT-PCR) amplification are routinely performed on paraffin-embedded tissue samples. Methods such as Southern or northern blot, requiring large fragments of DNA, are less likely to work consistently with fixed tissues.

Collection Tubes for Molecular Testing

Phlebotomy collection tubes are available with a number of different additives designed for various types of clinical tests. A selection of collection tubes commonly used for molecular biology studies is listed in Table 15.1. Some anticoagulants used in blood and bone marrow collection may adversely affect analytical results. Heparin has been shown to inhibit enzymes used in molecular analysis, such as reverse transcriptases and DNA polymerases in vitro.[1,2] The influence of this inhibition on molecular analysis is commonly accepted; however, heparinized samples have been processed successfully in many laboratories, and acid citrate dextrose (ACD) tubes are also used routinely for molecular tests.

TABLE 15.1 A Selection of Collection Tubes

Additive	Color	Nucleic Acid Testing
None	Red	Chemistry, serum, viral antibody studies
Sodium heparin (freeze-dried)	Green	Immunology, virology studies
Sodium heparin	Brown	Cytogenetic studies, molecular studies
Tripotassium EDTA (7.5%–15% solution)	Lavender	Virology, molecular biology studies
Acid citrate dextrose solution	Yellow	Molecular biology studies

The various experiences with heparin may reflect levels of resistance of enzymes from different sources. The assay design also has an effect. For instance, the inhibition of DNA polymerases compromises amplification of larger PCR products more than short ones. Due to the possible effects of heparin, trisodium ethylenediaminetetraacetic acid (EDTA; lavender-top) or ACD (yellow-top) tubes are recommended for most nucleic acid assays involving enzymatic treatment of the sample nucleic acid (Fig. 15.3).[3] High levels of disodium EDTA (royal blue–capped tubes used for trace element studies) may also inhibit enzyme activity and should be avoided. One of the advantages of signal amplification methods such as bDNA technology is their decreased susceptibility to the chemical effects of anticoagulants. When a specimen is received in the laboratory, the anticoagulant is not usually noted in the accompanying documentation; the technologist should be aware of the type of collection tube used, especially if a specimen is received in heparin or other additive that may affect the test results.

Tissue processing for molecular analysis requires caution to avoid contamination, especially where PCR or other amplification-based tests are to be performed. Recommendations for tissue processing for molecular diagnostic testing may be found in CLSI document MM13A, "Collection, Transport, Preparation, and Storage of Specimens for Molecular Methods; Approved Guideline."

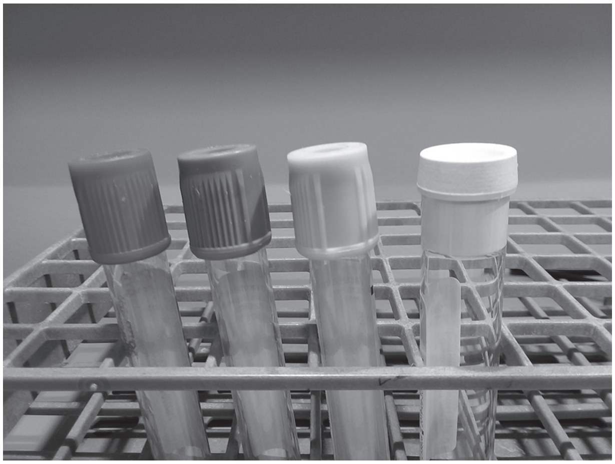


FIGURE 15.3 For molecular analysis, blood or bone marrow specimens collected in EDTA *(lavender-cap)* or ACD *(yellow-cap)* tubes are preferred. Heparin *(green cap)* is used for cytogenetic tests. Immunoassays or mass-spectrometry methods may be performed on serum collected in tubes without coagulant *(red cap tubes)*. See Color Plate 11.

In addition to the standard collection tubes, special collection tubes are designed particularly for stabilization of nucleic acids for molecular testing. These include the plasma preparation tubes (Vacutainer PPT, Becton Dickinson), which contain, in addition to the standard anticoagulants, a polymer gel that separates granulocytes and some lymphocytes from erythrocytes upon centrifugation. This type of tube is used for HIV and hepatitis C virus (HCV) analysis. The PPT tube is also used for separation of white cells for genetics, flow cytometry, and engraftment testing. The tube stoppers have a black "tiger" pattern over the appropriate anticoagulant color designation.

Specialty collection tubes are designed to stabilize RNA. These tubes contain proprietary RNA stabilization agents that maintain the integrity of the RNA from collection through isolation.[4-6] Separated white blood cells from standard collection tubes can also be lysed in phenol-containing solvents and the lysate stored at –70°C to stabilize RNA for several days.

Stabilization of transcripts has become increasingly important with the increased use of methods measuring RNA transcript levels. For example, quantitative RT-PCR analyses rely on transcript number to monitor the level of tumor cells or microorganisms.

Serial analysis requires that the specimens received in the laboratory at different times be handled as consistently as possible. Immediate stabilization of the RNA is important for consistent results. The laboratory procedure should include pre-analytical methods for specimen preservation and storage conditions. There are a variety of published methods for nucleic acid isolation; some accompany reagent sets used for testing of particular analytes. Isolated RNA quantity and quality can be measured where appropriate. Spectrophotometry is usually used for this purpose; however, fluorometry and gel electrophoresis are also used.

An alternative to standard collection tubes is the use of impregnated paper matrices for nucleic acid preservation and storage.[7,8] Specimens stored in this way are appropriate for amplification procedures (PCR) but not for assays requiring larger amounts of nucleic acids.

Precautions

Standard precautions are recommended by the Centers for Disease Control and Prevention for handling potentially infectious specimens. All specimens are potentially infectious, so they should all be handled with standard precautions using proper personal protective equipment (PPE) to prevent disease transmission. **Transmission-based precautions** including respirators are used with airborne or contact-transmissible agents. **Contact precautions** are designed for direct patient care where there is the potential for direct exposure to infectious agents on or from the patient. In general, standard precautions, including gloves and gowns as PPE, are used by the molecular laboratory technologist who has no direct contact with patients. Eye protection or masks are required in cases where frozen tissue is being processed or where spraying or splashing of a sample may occur.

Gloves are important, not only as part of standard precautions but also to protect nucleic acids from nuclease degradation (Fig. 15.4). Gloves are absolutely required for handling of RNA (Fig. 15.5). DNA is less susceptible to degradation from contaminating DNases; however, repeated handling of samples without gloves will adversely affect the integrity of the DNA over time. Standards and controls that are handled repeatedly are the most likely to be affected. Originally, having separate areas for DNA and RNA isolation was recommended. Some laboratories, however, isolate DNA and RNA on the same bench space. This requires maintenance of RNase-free conditions for all nucleic acid isolation. In a common isolation arrangement such as this, care must be taken that no inappropriate enzyme exposures occur if DNA samples are treated with RNase or if RNA samples are treated with DNase.

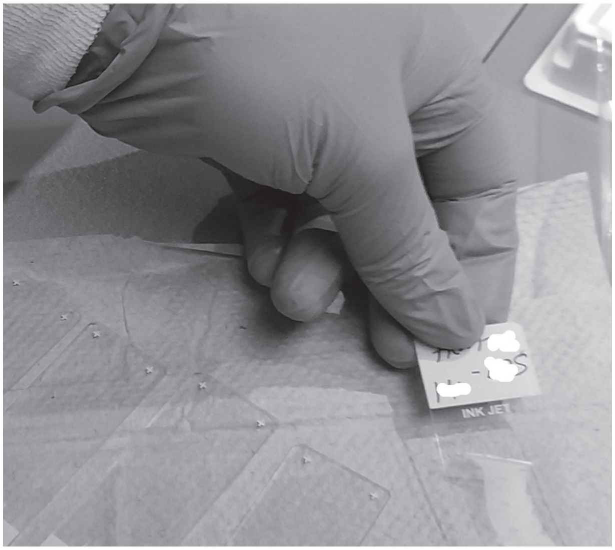


FIGURE 15.4 Handling specimens with gloves is recommended to protect the technologist and to protect the sample nucleic acids from nucleases and other contaminants.

FIGURE 15.5 Gloves are required for working with RNA.

Holding and Storage Requirements

Methods such as interphase and metaphase fluorescent in situ hybridization (FISH) and karyotyping require intact cellular structures or culture of cells so that only fresh specimens are acceptable. Many molecular methods, however, do not require the integrity of the cellular structure of tissue, only that the nucleic acid remains intact. In either case, circumstances may arise that require holding of specimens before analysis. DNA and RNA are stable when samples are collected and held under the proper conditions. For example, multiple tests may be performed on snap-frozen tissue specimens held at –70°C. Although most amplification methods are capable of successful analysis of limiting and challenging specimens, Southern or northern blot methods will not work consistently on improperly handled specimens or nucleic acids stored under less-than-optimal conditions. The College of American Pathologists (CAP), which provides accreditation standards that are followed by CAP-accredited laboratories to improve the quality of their testing, has published recommendations for sample and isolated nucleic acid storage (Table 15.2).

For long-term storage, isolated nucleic acid is preferred. Conditions for the storage of isolated nucleic acid have been recommended (Table 15.3). **Cryotubes** and specially designed labels are available for long-term nucleic acid storage at ultralow temperatures (Fig. 15.6).

The general rules for holding specimens for processing differ depending on the analyte and its stability in the cell. Written procedures indicate the proper handling of specimens for optimal performance of that procedure. For example, room-temperature storage is recommended for viral RNA in whole blood. Refrigeration of blood will cause neutrophils to degranulate, releasing enzymes that affect free virus particles. On the other hand, specimens in EDTA tubes held at 4ºC may not show significant loss of human RNA stability, depending on the gene. Still, blood or bone marrow for human RNA testing should be processed within 24 hours. Blood and bone marrow specimens sent to outside laboratories for molecular analysis can be shipped overnight at room temperature or with ice packs. Tissue is best shipped frozen on dry ice.

Isolated DNA of sufficient purity can be stored at room temperature for several months or at least 1 year in the refrigerator. Purified DNA can be stored at freezer temperatures (–20°C to –70°C) in tightly sealed tubes for up to 10 years or longer. Freezer temperatures are preferred for long-term storage; however, a clean DNA preparation in frequent use is better stored in the refrigerator so as to avoid DNA damage caused by multiple cycles of freezing/thawing. Shearing of DNA by freeze/thawing cycles can also occur in a frost-free freezer. As previously stated, PCR and methods that do not require large intact fragments of DNA are more forgiving with regard to the condition of the DNA.

Storage of isolated RNA at room temperature or refrigerator temperature is not recommended in the absence of stabilization. RNA suspended in ethanol can be stored at –20°C for several months. Long-term storage is best in ethanol at –70°C, although RNA suspended in diethylpyrocarbonate-treated water is stable for at least 1 month. As with DNA, the long-term survival of the RNA depends on the quality of the initial isolation and handling of the specimen.

TEST PERFORMANCE

Various commercial reagent sets or “kits” have become available for patient testing. Many of these reagent sets are approved by the Food and Drug Administration (FDA-approved) for use in clinical testing. The FDA-approved tests are considered “waived” tests by the CAP and other agencies charged with inspection of medical laboratories. For these tests, most of the calibration, validation, and other quality assurance measures have been established by the manufacturer and approved by the FDA. If the test is performed strictly according to the manufacturer’s protocol, implementation of the test only requires verification of the accuracy, precision, reportable range, and reference range. If the laboratory modifies the protocol, for instance, doing an amplification step in a thermal cycler model different from that used to validate the test or performing the test on blood collected in a different anticoagulant, then the laboratory must establish clinical performance.

Tests in the molecular laboratory that are individually designed or are adaptations based on published methods are called **laboratory-developed tests (LDTs)** and, like modified FDA-approved tests, are considered non-waived for the purposes of laboratory review.

TABLE 15.2 Specimen Storage Requirements for Nucleic Acid Extraction[‖]

Nucleic Acid	Sample	Temperature	Time
DNA	Whole blood, buffy coat, bone marrow, fluids	22°–25°C	24 hours
		2°–8°C	72 hours*
		−20°C	At least 1 year*
		−70°C	More than 1 year*
	Tissue	22°–25°C	Not recommended
		2°–8°C	Up to 24 hours[†]
		−20°C	At least 2 weeks
		−70°C	At least 2 years
	Microorganisms in culture	22°–25°C	24 hours[‡]
		2°–8°C	72 hours[‡]
		−20°C	2–4 weeks[‡]
		−70°C	More than 1 year
	Cell lysates in GITC	22°–25°C	1–2 weeks[§]
RNA	Whole blood, buffy coat, bone marrow, fluids	22°–25°C	Not recommended[§]
		2°–8°C	2–4 hours*
		−20°C	2–4 weeks*
		−70°C	More than 1 year*
	Fluids collected in specialty	22°–25°C	5 days
	RNA protection tubes	2°–8°C	7 days
		−20°C	2–4 weeks
		−70°C	At least 7 months
	Tissue	22°–25°C	Not recommended
		2°–8°C	Not recommended
		−20°C	Not recommended
		−70°C	At least 2 years
		−140°C (nitrogen vapor)	At least 2 years
	Cell lysates in GITC[¶]	22°–25°C	1–2 weeks[§]
	Cell lysates in RNA storage solution (Ambion)	22°–25°C	1 week
		2°–8°C	1 month
		−20°C	More than 1 year
	Microorganisms in culture	22°–25°C	24 hours[‡]
		2°–8°C	72 hours[‡]
		−20°C	2–4 weeks[‡]
		−70°C	More than 1 year

*Separation of white blood cells is recommended to avoid hemoglobin released upon hemolysis of red blood cells.

[†] Tissue types differ in stability and nuclease content.

[‡] Nucleic acid from cultured organisms is best isolated immediately on harvesting fresh cultures.

[§] RNA status depends on the type of cell or tissue and the gene under study.

[‖] Depending on gene expression, adequate RNA may be isolated within a few hours. Storage of cell lysates in a stabilizing buffer is best for maintaining RNA.

[¶] GITC = guanidine isothiocyanate.

TABLE 15.3 Nucleic Acid Storage Requirements

Nucleic Acid	Matrix	Temperature	Time
DNA	TE* buffer or DNase-free water	22°–25°C	Up to 4 months
	Freeze-dried or dried on collection paper	22°–25°C	More than 15 years
	TE buffer or DNase-free water	2°–8°C	1–3 years[†]
	TE buffer or DNase-free water	−20°C	At least 7 years
	TE buffer or DNase-free water	−70°C	More than 7 years
RNA	TE buffer or RNase-free (DEPC[‡]-treated) water	22°–25°C	Not recommended
	TE buffer or RNase-free (DEPC-treated) water	2°–8°C	Not recommended
	TE buffer or RNase-free (DEPC-treated) water	−20°C	Up to 1 month
	RNA storage solution (Ambion)[§]	−20°C	More than 1 month
	Ethanol	−20°C	More than 6 months
	TE buffer or RNase-free (DEPC-treated) water	−70°C	Up to 30 days
	Ethanol	−70°C	More than 6 months

*10 mM Tris, 1 mM EDTA, pH 8.0.

[†] One year for Southern blot.

[‡] DEPC = diethyl pyrocarbonate.

[§] 1 mM sodium citrate, pH 6.4.

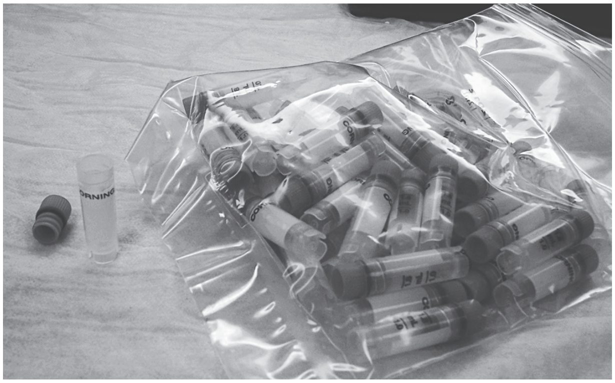


FIGURE 15.6 Cryotubes with tight-fitting lids are recommended for long-term freezer storage of DNA and RNA.

Development of new tests in the clinical laboratory requires validation of the performance of the method and reagents in accurately detecting or measuring the analyte.[9]

Clinical test parameters must be established from clinical trial data or published population studies. The positive predictive value and negative predictive value of the analyte status in a given disease state justify the use of the test for patients. The positive predictive value is the degree that a particular test result corresponds with the presence of a clinical state, such as disease symptoms or response to a therapeutic agent. The negative predictive value is the clinical specificity of the test, that is, the absence of that test result in the absence of the disease state or response.

Analytical test performance is assessed by several criteria (Table 15.4). Analytical test accuracy is a function of the sensitivity and specificity of the test.

TABLE 15.4 Measurements of Test Performance

Criteria	Definition	Example
Analytic sensitivity	Change in assay response with corresponding change in analyte	All positive reference standards tested positive with the new assay. The analytical sensitivity of the assay is 100%.
Clinical sensitivity	Ability of test result to predict a clinical condition	Of 100 patients with a gene mutation, 95 have a disease state, a clinical sensitivity of 95%.
Detection limit, limit of detection	Least detectable presence of the analyte	The t(14;18) translocation test can detect 1 translocated cell in 10,000 normal cells, a detection limit of 0.01%.
Analytic specificity	Ability to detect only the analyte and not nonspecific targets	The Invader assay for factor V Leiden successfully detected mutations in 18 positive specimens while yielding negative results for 30 normal specimens (no false positives).
Clinical specificity	Disease-associated results only in patients who actually have the disease conditions	Of 100 normal specimens, 1 displayed a gene mutation (1 false positive), a clinical specificity of 99%.
Precision	Agreement between independent test results	A quantitative method yields 99 results less than 1 standard deviation apart in 100 runs, a precision of 99%.
Reproducibility	Consistency of test results produced from the same procedure	A qualitative method yields 100 positive results when performed in 10 independent laboratories, a reproducibility of 100%.
Analyte measurement range	The range within which a specimen may be measured directly (without dilution or concentration)	A qPCR HSV assay yields reproducible linear results from 10 to 107 copies of HSV per 20 µL of CSF. Specimens within this range are measured directly.
Reportable range	The range within which test results are considered to be valid (with or without dilution)	A qPCR HCV assay yields reproducible results from 15–100,000,000 IU/mL of blood, requiring dilution for levels greater than 100,000 IU. Specimens within this range are reportable.
Reference range	Expected analyte frequency or levels from a population of individuals	The reference range for prostatic specific antigen is 0–4 ng/mL.
Analytic accuracy	Production of correct results	Of 100 specimens with mutations in the *HCM* gene, 99 are detected by sequencing, with no mutations detected in normal specimens.
Linearity	Quantitative correlation between test result and actual amount of analyte	A graph of test procedure results versus input analyte yields a straight line.

The **analytical sensitivity** of an assay equals

$$\frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN} \times 100}$$

The **analytical specificity** of an assay equals

$$\frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP} \times 100}$$

The **analytical accuracy** of an assay equals

$$\frac{\mathrm{TN} + \mathrm{TP}}{\mathrm{TN} + \mathrm{TP} + \mathrm{FN} + \mathrm{FP} \times 100}$$

where TN = true negative, TP = true positive, FN = false negative, and FP = false positive.

True positive and true negative are determined using a gold standard, such as an established reference assay or a definitive clinical diagnosis. Note that the clinical sensitivity is different from the analytical sensitivity, which may also be reported (Table 15.4). The analytical sensitivity is defined as the ability of the assay to detect the presence of the analyte, with changes in the assay response corresponding to changes in the amount of analyte present. Analytical sensitivity can be further refined by determining the detection limit or lower limit of detection. The detection limit is the lowest level of analyte that is consistently detected by the assay. In qualitative assays, the analytical sensitivity and the limit of detection are practically equivalent. In quantitative assays, the analytical sensitivity is defined by differences in quantitative response with quantitative differences in the reference analyte. For tests with quantitative results reported qualitatively (e.g., positive/negative) a cut-point quantity must be established, above which is considered positive and below which is considered negative. Thus, the quantity may be determined empirically or statistically using **probit analysis.**[10,11] These criteria, along with the analytical measurement range (AMR; Table 15.4), are documented as part of the test validation process.

The AMR is established by showing accurate test results on a range of values or test results. The validation should include sufficient validation samples at AMR levels in order to demonstrate performance. For qualitative results, such as positive/negative or homozygous wild-type, homozygous variant, or heterozygous, sufficient numbers of each type of sample must be available. For array analyses or next-generation sequencing, all possible alleles do not have to be verified; however, the ability to accurately detect an adequate sampling of alleles in a range of concentrations (for somatic variants) should be established. Actions taken for samples above or below the AMR should be documented in the validation record. Such results may be reported as "greater than" or "less than" the limits of the AMR, or the validation may include methods and criteria for dilution or concentration of samples. Note that dilution and concentration do not increase the AMR.

Other criteria for validation include precision and reproducibility. Precision is the correct qualitative result or the degree of variation in a test of quantitative results when performed on the same sample multiple times. Precision should be established throughout the reportable range. Reproducibility is the ability of the test to resist variations in testing conditions, such as reagent lots, test days, different technologists, or other sources of random error. A very precise test on one instrument in one lab on a particular day may not be reproducible when performed on a different instrument or with a different test lot.

The purpose of test validation is to demonstrate that a procedure is ready to be implemented as a clinical test.[12] Test validation is performed on specimens of the types that will be encountered in the routine use of the test, such as frozen tissue, paraffin-embedded tissue, body fluids, and cultured cells. The number of specimens tested varies with the procedure and the availability of test material. Archived specimens are often used for this purpose. The results from the new test are compared with those of established procedures that may have been performed on these specimens or with the clinical diagnosis. A standard form may be designed for the preparation of reaction mixes by the test parameters determined in the validation process. Laboratory information systems are also programmed to produce setup forms for validation and routine patient testing.

Predeveloped and FDA-approved molecular methods are increasingly available. According to the FDA, an approved test is one that has been accepted by the FDA as safe and effective for its intended use, based on the manufacturer's data. The manufacturer conducts studies to show that the test performs as claimed and does not present an unreasonable risk. The test is offered for sale after a premarket-approval application has been reviewed and approved by the FDA. A test that has been cleared by the FDA has been shown to be similar to other tests already marketed, based on the manufacturer's data. For such a test, the manufacturer submits comparison results in a "premarket notification" that the FDA reviews and determines that the tests are substantially equivalent.

When FDA-approved or FDA-cleared methods are incorporated, the test performance is verified by using the purchased reagent sets to test validation specimens. This verification establishes that the results of the commercial test performed in the individual laboratory are as predicted by the developer. If the commercial test is modified, validation is required to show equal or superior performance of the modified procedure.

Once a procedure has been established, the method is documented in the laboratory according to CLSI guidelines. The procedure description should include detailed information, for instance, primer and probe sequences, their purification conditions, and labeling. A copy of the standard form used to set up reaction mixes is included in the procedure description. A clear description of formulas and reporting units is required for quantitative results. Interpretation of qualitative data, acceptable ranges (e.g., band patterns), product sizes, melting temperatures, and reasons for rejecting results are required information. Methods used to score FISH or array results relative to internal control loci are also part of the written procedure. It is useful to incorporate pictures of gel patterns or instrument output data showing positive, negative, heterozygous, or other reportable results.

The clinical accuracy of a test is determined by correlation of the test results with morphological, radiographic, or other clinical data. For rare conditions or where clinical associations are well established, literature reports of correlation with the analyte state may be used. The indications for ordering the test are established based on the clinical utility as determined by the validation process and are documented in the procedure manual. For forensic testing, all aspects of the test, from validation to test reporting, should adhere to guidelines established by the DNA Advisory Board Standards and the Scientific Working Group on DNA Analysis Methods.

The procedure manual or standard operating procedure is maintained in the laboratory and reviewed at least annually. If a test is discontinued, the written procedure, noted with the dates of initial use and retirement, is kept for at least 2 years. Some laboratory professionals maintain retired procedures for longer periods.

Next-Generation Sequencing

Validation parameters for NGS will vary between tests and typically include a description of the analytical target (exons, genes, or targeted regions, such as introns or promoter sites) and bioinformatics used for analysis. Sample pooling methods are designed such that individual sample identity is maintained throughout testing and analysis. Criteria and thresholds for variant calling include minimum read coverage depth, base or variant quality scores, and allelic read percentage. These criteria may be different based on application (e.g., detection of germline versus somatic mutations). If there are target regions where variant calls can be affected by sequence structure, such as highly homologous regions or long nucleotide repeats, the laboratory must document how interference is mitigated or avoided.

Performance characteristics, including sensitivity, specificity, and precision (reproducibility), for variants cover examples of the different variant types (single-nucleotide variants, indels, copy-number and other structural variants) to be detected. The limits of detection for variants in samples are established for heterogeneous genotypes (e.g., heterogeneous tumor samples or microchimerism). For somatic analyses of solid tumors, the sensitivity and limit of detection will be affected by the tumor cell percentages in the tested tissue. This information is also documented.

The detection rate of variants by bioinformatics processes and software is applicable to exome and genome sequencing data analysis for suspected genetic disease components. Acceptance and rejection criteria for the results generated by the analytical bioinformatics process are based on quality control (QC) parameters established during instrument and software optimization, including base and mapping quality scores, read coverage, and numbers and types of variants.

The type and quality of sequence databases used for annotation (identification of the significance of variants) are established as part of the validation process. Databases may be constitutional for genetic diseases, population based to identify the frequencies of minor alleles, or cancer specific. Clinical laboratories may establish internal databases as well.

Calibrators and Method Calibration

Calibration is the establishment of conditions for instrument response/method result association with the true value of quantitative analyte measurements. These values are represented in calibrators. Calibrators are samples of known amounts/concentrations of molecules of the same type measured in the assay, such as an admixture of PCR products in known proportion or an RNA transcript calibrator in a matrix of normal RNA. Calibrators can be purchased from manufacturers. Some are supplied with instruments for instrument maintenance. Others are supplied with reagent sets for validation or re-validation purposes. Patient specimens previously tested by a validated method or specimens used in proficiency testing are also calibrators. When a calibrator is in a different matrix (diluents or mixture) than the samples to be tested, such as with general-purpose commercial calibration sets, commutability (lack of matrix effects) must be established.[13]

Calibrators covering the AMR demonstrate linearity if calibrator values are plotted against instrument or test output values. In the test validation, the AMR is established with samples of the appropriate type and matrix. Calibrators may not necessarily be supplied in the same matrix but can be diluted or dissolved into the appropriate matrix. The source, quality, and preparation of the calibrators should be documented. Certificates of quality from the vendor accompany commercially supplied calibrators. Re-calibration is performed if calibration does not meet the required standards of linearity or accuracy.

Calibration verification should be done at regular intervals, at least every 6 months, or at intervals established by the laboratory. Calibration verification may be required upon changes of or major repairs to instruments or changes in reagent lots that might affect test performance or when quality control indicates shifts or unacceptable errors in test control results. Verification should demonstrate the continued linearity of correlation between the calibrator values and test results.

Controls

Controls are samples of known type or amount that are treated like and run with patient specimens. Interpretation of test results always includes inspection of controls and standards to verify acceptable test performance. With qualitative tests, a positive, a negative, and, in some cases, a **sensitivity control** are required. The sensitivity control defines the lower limit of detection for more meaningful interpretation of negative results. For nucleic acid amplification techniques, an **amplification control** is used. The amplification control is a target that should always amplify. The amplification control distinguishes true-negative amplification results from false negatives resulting from amplification failure. In quantitative methods, high-positive, low-positive, and negative controls are included with each run. The high and low levels should be similar to critical points in the assay, such as the lowest detectable level of analyte. An acceptable range for high- and low-positive controls may be established by repeated runs of each control sample on different days by different personnel. The average of ±2 or ±3 standard deviations defines the acceptable range for positive controls. A set of controls may also be included with runs in the form of a calibration curve with each run.

Real-time PCR methods that automatically determine analyte levels require measurement of a **standard curve** or dilution series of analyte levels encompassing the levels expected from the patient specimens. On some instruments, the standard curve must be run simultaneously with the specimens, whereas in others, previously determined curves may be loaded into the software. Alternatively, results can be calculated manually by linear regression of the test results, using standard-curve data in spreadsheet software.

In methods requiring detection of a target-specific product or relative amounts of target, **internal controls** are run in the same reaction mix as the test specimen. For example, RNA from housekeeping genes is used for internal controls in methods quantifying infectious agents or detecting tumor cells by expression of tumor-specific transcripts.[14] Centromere-specific probes serve as internal controls in FISH analyses, as do housekeeping gene probes on microarrays. The presence of an internal control supplies a base for normalization of results. In PCR, the internal control distinguishes false-negative results from failed amplifications. Internal controls that are amplified in the same tube with sample templates are designed to not interfere with or inhibit target amplification, which could yield a false-negative result. Failed internal controls are documented and call for a repeat of the assay.

The controls and standard curve should cover the critical detection levels or results of the method. Control results are continually monitored to spot trends or spikes outside of tolerance limits. Coefficients of variance or standard deviations of quantitative control levels should also be calculated at regular intervals. Laboratory professionals may establish criteria for control tolerance limits and document actions to be taken in the event of an unacceptable control result.

Control probes for arrays and in situ hybridization methods depend on the signal pattern to be tested (amplification, structural change, or deletion) and the tissue being tested. Internal control probes may map to the centromere or other locus close to the test probe–binding site. External controls, probes run on specimens in parallel to the test specimen, are used for loci that may not have an internal control, such as the Y chromosome in females. Controls for most assays are commercially available or supplied with reagent sets. When commercially prepared controls are not available, alternative controls (previously tested samples tested in duplicate, samples tested by a different method, or samples tested in independent laboratories) are used. Nucleic acid from these controls is best prepared in larger quantities, aliquoted, and stored in conditions where they are most stable. Just as with new lots of other reagents, new aliquots are tested with old aliquots to verify consistent control results, and the procedure for the use and storage of all controls is documented.

QUALITY CONTROL

Positive, sensitivity, and negative controls should be included in each run of patient samples. For tests with multiple targets, controls can be systematically rotated for different targets in each run. Controls for quantitative tests should reflect the clinically critical levels of test results. Alternatively, the laboratory can establish an individualized quality control plan (IQCP). This plan includes monitoring of the extraction and amplification steps of the assay. Tests with electronic or built-in controls are appropriate for the IQCP.

Tolerance or acceptability levels of control values should be established based on the validated precision and reproducibility of the assay. Monitoring can be performed using Levey–Jennings or modified Levey–Jennings plots[15] or cusum plots,[16] and tolerance limits can be expressed as Westgard rules.[17] The results of these monitors are reviewed before reporting test results. Review takes place at least monthly and may include multiple systems and networked laboratories.[18] If the control results violate the set limits, then corrective action (repeating the run with replacement controls) has to be taken and documented. If patient specimen availability precludes repeating a run, further investigation into the nature of the control failure and its effect on the patient result may allow acceptance of the sample results without repeating. Comparison of test values with previous results or historical averages may serve to increase confidence in the test results.

QUALITY ASSURANCE

Monitored controls and conditions are reviewed monthly or at determined intervals to document outliers or systemic error. These monitors can be established by the IQCP. Periodic review and documentation of test results are required for all clinical testing, including molecular tests. Review might be, for example, in the form of rates of positive and negative results compared with expected rates from independent sources, such as published results, over time. This type of monitoring reveals trends or shifts in the rates of positive or negative results. Critical values that require physician notification are established by validation and confirmed by monitoring.

As with other types of quantitative testing, defined dynamic ranges, sensitivity levels, and accuracies are subject to review. For instance, a test that determines viral types by melt curves must include a defined narrow temperature range for the T_m of each viral type (Fig. 15.7). These values are established during the test validation and should be reviewed periodically using high, low, sensitivity, and negative standards in each run. Band patterns, melt curves, and peak characteristics should be defined with regard to how the results of the test are to be interpreted.

Assay levels that distinguish positive from negative results (**cut-off values** or cut points) must be well defined and verified at regular intervals. In assays such as single-strand conformation polymorphism, in which control patterns are not identical from run to run, normal

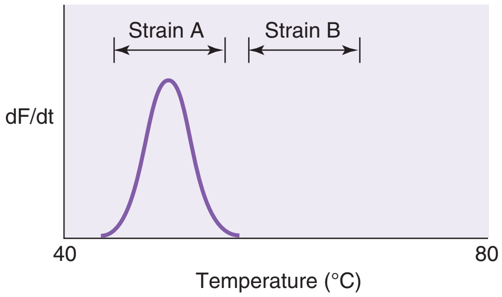


FIGURE 15.7 T_m ranges defined for two strains (A and B) of a theoretical microorganism are nonoverlapping. The melt curve shown would indicate that the test specimen contains strain A.

controls for each scanned region are included in every run. Quantitative results should be within the linear range of the assay. The linear range is established by measuring dilutions or known concentrations of standard and establishing a direct correlation (standard curve) between test output and standard input. The technologist may observe that raw data are consistent with the final interpretation of the results. For example, if a viral load is interpreted as negative, the raw data should be below the cut-off value established for the test. Calculations and comparisons with standards used to verify test results should be described in the laboratory procedure manual.

For sequencing procedures, information about the genes tested and variants found is continually updated, along with the databases used for annotation. Documentation of the method and results of verification of genetic or germline variants found is maintained by the laboratory.

When controls or results exceed acceptable limits, corrective action is taken. Corrective action can be minor adjustments or a repeat of a sample or run or more extensive changes to a system. Documentation of what is done in corrective action reports is an important component of quality assurance.

Quality assurance also includes documentation and maintenance of laboratory procedures and methods. These procedures are frequently updated with events ranging from major advances in instrument technology to changes in reagent availability. Electronic document control systems facilitate editing of documents as well as communication of the status of policies and procedures with laboratory personnel.

Test results may show discrepancies with other laboratory findings, with clinical observations, or with the laboratory's own preliminary results. These discrepancies are documented along with any corrective action taken, if necessary. Due to the nature of molecular pathology testing (increased sensitivity, high resolution, varied methodologies used) discrepancies may be explained by the technical aspect of the test. In these cases, the test validation should include data on the clinical significance of the analyte as detected or measured by the molecular test, for example, detection of analyte below levels of clinical significance. Statistics on these results may also be collected as part of quality assurance and control.

INSTRUMENT MAINTENANCE

Instruments used in the molecular laboratory must be monitored and maintained for consistent performance and accurate test results. Manufacturers supply recommendations for routine maintenance. Service contracts are used to provide support from trained service technicians. Laboratory professionals maintain a schedule and instructions for all routine maintenance, such as checking temperature settings, timers, and signal background levels. Parts are replaced as required or specified by the instrument manufacturer. Maintenance schedules should reflect the amount of use of the instrument. Most routine maintenance and minor troubleshooting, such as replacing bulbs or batteries, may be performed by the technologist with the aid of clear instructions from the manufacturer or as prepared by laboratory management. These instructions must be readily available to the technologist in the event of instrument malfunction. Technologists should be aware of the limits of user-recommended repairs and when service calls are indicated (Fig. 15.8). Laboratory professionals document all maintenance, service calls, calibrations, and part replacements.

Refrigerators and freezers used to store patient material and reagents are monitored at least daily (Fig. 15.9). Maximum/minimum thermometers register the highest and lowest temperature reached between monitoring

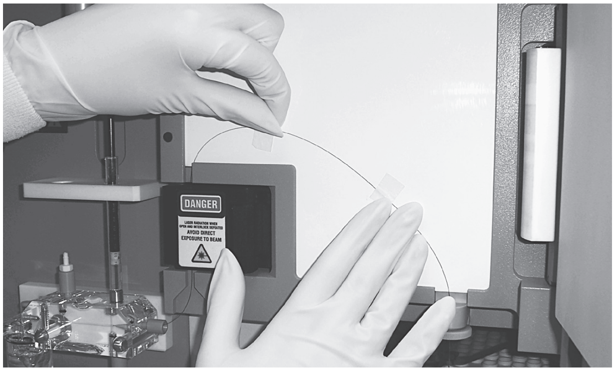


FIGURE 15.8 Routine maintenance, such as capillary replacement and instrument cleaning, is performed by the laboratory technologist. Dangerous or complex maintenance, such as repair or replacement of a laser source, is performed by the service representatives.

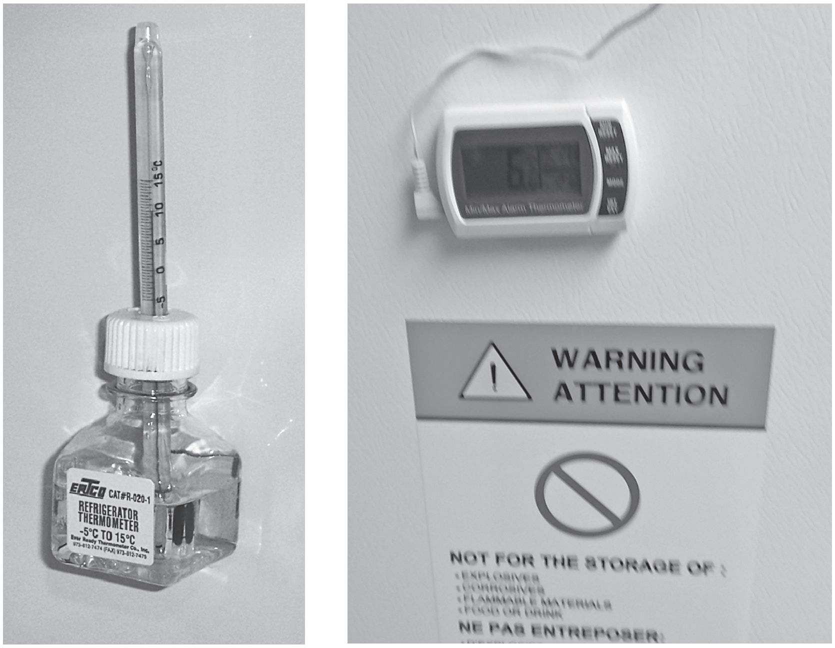


FIGURE 15.9 Certified chamber thermometers are used for monitoring temperatures in incubators, refrigerators, ovens, and freezers. Thermometers must be checked at least annually by comparison with a NIST-calibrated thermometer. "High-low" thermometers or commercial monitoring systems are used when monitoring is required when the laboratory is not in active operation.

points. During a busy shift, refrigerators and freezers may be opened frequently, causing the temperature to increase temporarily. This must be taken into account while monitoring. Out-of-range temperatures (e.g., more than ±2°C of the set temperature) are recorded. Confirmatory temperature checks at different times during the shift are required before further action is taken. Heat blocks, incubators, ovens, and water baths are also monitored for temperature stability and accuracy. U.S. National Institute of Standards and Technology (NIST)–certified standard thermometers are used for this type of monitoring. Thermometers verified by a NIST thermometer are also acceptable for this purpose. Specialized dry bath thermometers, which are encased in a microfuge tube of mineral oil, or electronic temperature probes are used to monitor heat blocks. All storage and incubation equipment is kept clean and free of contaminated specimens and expired reagents. If out-of-date reagents are used for research or other purposes, they should be well marked and/or maintained in a separate area from the clinical test reagents. Although standard thermal cyclers have no automated moving parts, they may decline in temperature control over time. This is especially true of block cyclers where hot and cold spots develop within the sample block. For this reason, thermal cyclers are checked periodically for proper temperature control. Thermometers with flexible probes (type K thermocouples) are convenient for checking representative wells in a block thermal cycler (Fig. 15.10). Approaches differ regarding whether each well should be checked with each routine measurement or whether representative wells should be checked, with different wells checked each time. In some laboratories, test reactions are run in different wells to demonstrate successful amplification at each position on the block. More thorough and accurate temperature measurements are achieved with computer systems, with fixed or flexible probes designed to measure the temperature in all wells throughout a PCR program, including ramping of the temperature up and down, overshooting set temperatures, and temperature drift during the hold phase of each step. Nonblock thermal cyclers, such as air-heated or modular instruments, are tested with probes modified to fit as the capillaries or tubes used in these instruments. Real-time thermal cyclers require additional maintenance of the detection system. Manufacturers supply materials to check for "bleed through" of fluorescence of different wavelengths. Background measurements are made using water or buffer samples. Each laboratory will establish the type and frequency of scheduled maintenance.

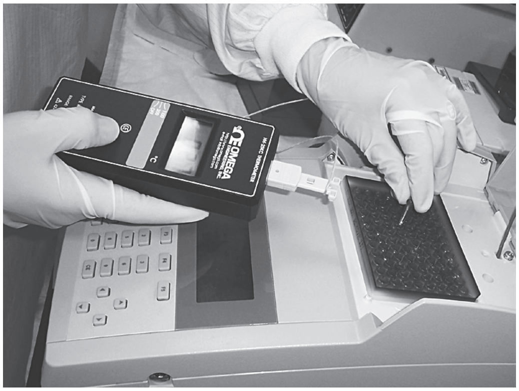
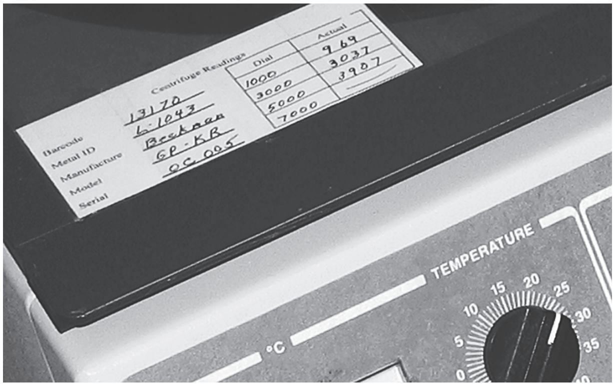


FIGURE 15.10 Block thermal cyclers may be monitored using a thermometer with a flexible probe. More thorough monitoring is performed with multiple temperature probes and software that follows the ramping temperatures as well as the holding temperatures.



Centrifuge Readings

Barcode 13170
Metal ID L-1043
Manufacture Beckman
Model GP-KR
Serial QC 005

Dial	Actual
1000	969
3000	3037
5000	3907
7000	

FIGURE 15.11 Centrifuge speeds are checked at least annually, and the results of actual and set speeds are posted on the instrument.

Centrifuges and microcentrifuge speed controls are monitored at least annually using a tachometer. In some institutions, technologists perform this calibration. Alternatively, an institutional engineering department may provide this service. The actual speed of rotation is determined and recorded, along with the set speed or setting number on the centrifuge. This information is then posted on the instrument (Fig. 15.11). Automatic pipettors used for dispensing specific quantities of reagents are checked for accuracy before use and at 6-month intervals or as required according to use. Gravimetric methods, where measured samples of water are pipetted to an analytical balance, have been used for many years. The weights are converted to volumes. The mean of several measurements from the same pipet reveals its accuracy. The standard deviation or **coefficient of variance** is calculated to determine the degree of reproducibility (imprecision) of the pipet. Service providers will clean and check pipets on a per-pipet charge.

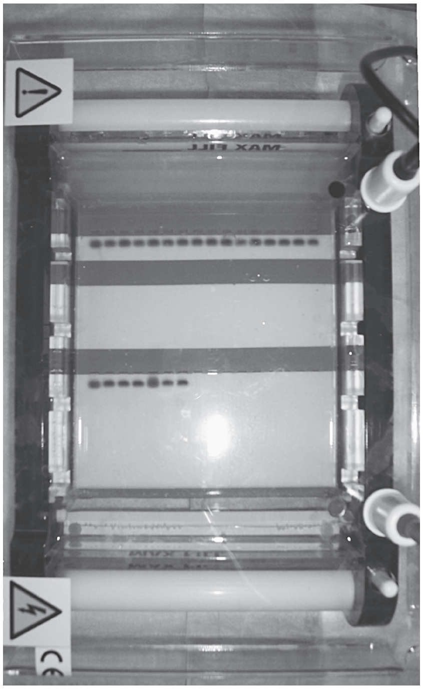


FIGURE 15.12 Gel electrophoresis equipment must be maintained free of precipitate and properly handled to avoid shock exposure.

Electrophoresis power supplies are tested at least annually to ensure delivery of accurate voltage and current. Personnel should be trained in safe operation. Leads and connectors to gel baths are kept free of precipitate (Fig. 15.12). Salt precipitation can be avoided by not leaving buffer in gel baths after electrophoresis runs. Capillary systems require cleaning of buffer and polymer delivery channels as well as replacement of the polymer at least twice per month or as required per its expiration date. Capillaries are also replaced according to their suggested life span in number of uses or time in use. Temperature-controlled electrophoresis equipment, such as capillary systems and those used for constant-temperature gel electrophoresis, is monitored for accurate temperature settings as recommended by the manufacturer.

Photographic equipment is frequently used in molecular laboratory procedures. Autoradiograms resulting from radioactive or chemiluminescent methods are developed in automated equipment or manually. The processing equipment is maintained with fresh fixing and developing solutions free of debris or sediment. Digital cameras are rapidly replacing Polaroid cameras for documenting gel data. Cameras should be firmly mounted and adjusted for optimal recording of gel data, free of shadows, dust, and other photographic artifacts (Fig. 15.13).

Periodically, background measurements are required on such instruments as fluorometric detectors (including real-time thermal cyclers), luminometers, and densitometers. Instrument manufacturers provide guidelines for acceptable background levels. Laboratory professionals may use these measurements or establish their own acceptable background levels. Corrective action, such as cleaning or filter adjustment, is documented in the laboratory maintenance records. Ultraviolet (UV) illuminators are kept free of dust and properly shielded while in use. The technologist keeps track of the life span of the UV light source and replaces it accordingly.

Signal response from spectrophotometers used to measure DNA, RNA, protein concentrations, and colorimetric assays and turbidity is checked annually or as recommended by the manufacturer. Maintenance also includes scanning through the range of wavelengths used (e.g., 200 to 800 nm) with supplied materials or filters. Operation manuals will include instructions for calibration and maintenance.

Fume hoods and laminar flow (biological safety) hoods (Fig. 15.14) are monitored annually for proper airflow. Fume-hood testing requires special equipment and is likely to be performed by building engineers.

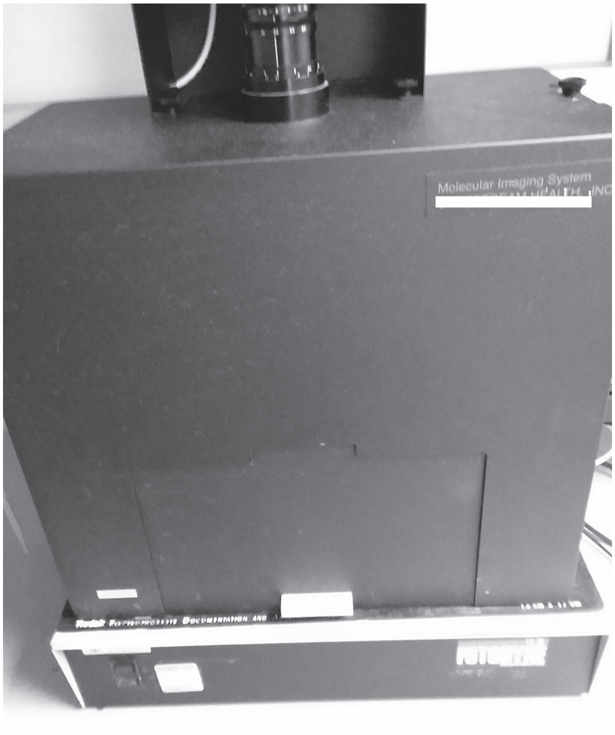


FIGURE 15.13 Cameras should be mounted securely for gel photography. A digital camera, shown here, is mounted on a photographic hood. A mask shields the ultraviolet light source except for the area where the gel is illuminated.

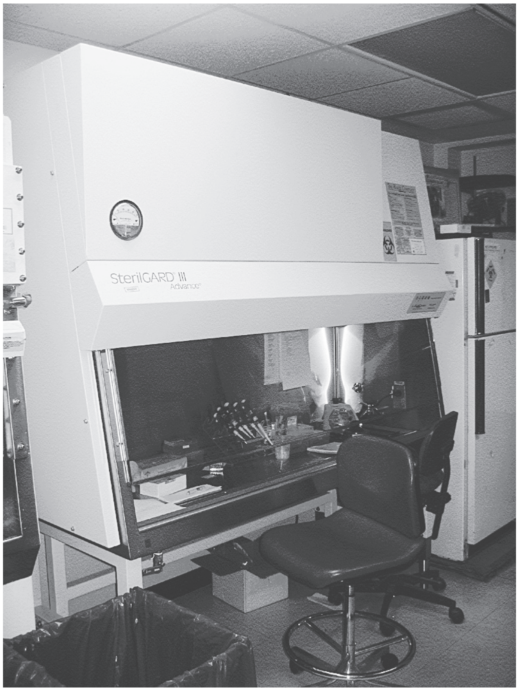


FIGURE 15.14 Laminar flow hoods are used to protect against biological hazards and to help maintain a sterile environment.

Laminar flow hoods are tested for proper filter performance and air displacement. This testing is performed at least annually or upon installation or movement of the hood. Professional service technicians or the hood manufacturers usually provide this type of certification.

For all detection systems, regular monitoring of functional characteristics will reveal any drift or trends that might affect test results. Tolerance limits should be established to warrant intervention by maintenance or recalibration of the instrument. Scheduled and unscheduled maintenance is documented and kept in the laboratory records. These records should be readily available to the technologists using the equipment.

Instrument Calibration

Calibration is fitting an instrument or test system output with the actual concentration of a reference analyte by testing and making appropriate adjustments. In **calibration verification,** materials of known concentration throughout the reportable range are tested as patient samples to ensure the test system is accurate. If calibration verification fails, recalibration is required. CLIA-88 regulations, 42 CFR 493.1255(b)(3), recommend the performance of calibration verification at least every 6 months or when major components, instrument software, or lots of reagents of the test system are altered. Recalibration is also required if proficiency or other quality control testing fails or in the event of major instrument malfunction and repair. Manufacturers of test systems may also provide calibration schedules and instructions on how to perform calibrations. Laboratory professionals must verify the calibration of systems performed by the manufacturer.

A variety of materials may be used for calibration, including previously tested specimens, reference standards, commercial calibrators, and proficiency testing material. There must be an independent assessment of the actual measurement of the calibration material. Once established, calibrator results should always be specified ranges of values. Calibration materials should cover at least three levels of measurement: low, medium, and high points. Analytes used for calibration should be in the same matrix (e.g., plasma or urine) as the patient specimens.[19] Calibrators are prepared and used separately from quality control standards (e.g., positive, negative, sensitivity controls) for routine runs.

REAGENTS

When reagents are replaced in a test method, the new lot is ideally tested on a previously positive and negative specimen as well as the run controls. Instructions on the preparation of reagents and the quantities used in each assay are included in the written laboratory protocol for each procedure. Lot numbers and working stocks of probes and primers used in amplification methods are documented and matched to test performance in the runs in which they were used. The sequences of primers and probes are also documented because any sequence errors made during ordering or synthesis of the primers will adversely affect the amplification specificity or even result in amplification failure. Probes used for linkage analysis and array technology are periodically updated as new markers are discovered, so the probe sequences used for a given test should be recorded.

Primers are a critical component of PCR procedures. Primers are most conveniently supplied in lyophilized (freeze-dried) form from the DNA synthesis facility (Fig. 15.15). The supplier will also provide information on the quality, method of purification, molecular weight,

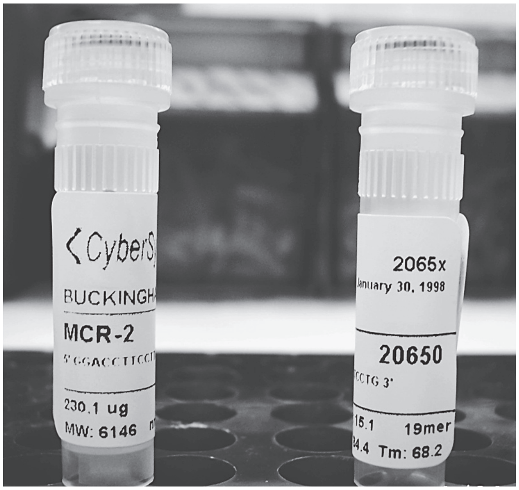


FIGURE 15.15 Primers are purchased from DNA synthesis facilities. On receipt in freeze-dried form, the primers are easily resuspended in nuclease-free water or buffer to make a stock solution. The stock solution is then diluted into working stocks.

and number of micrograms of dried primers. This information is used to rehydrate the primers to a stock solution concentration required for the PCR protocol. The resuspended primers are then diluted into working stocks.

Probes used for real-time PCR are supplied in solution or in dried form, for example, a 100-µM supplied or resuspended stock solution that is diluted to 4- or 5-µM working stock before use in the procedure. When new working stocks are prepared (diluted from the probe stocks or resuspended primers), they are treated as new reagent lots. Master mixes of primers, probe, buffer, nucleotides, and enzyme may be prepared or purchased and used as working stock.

As is required for all reagents, instructions on preparation of primers, probes, and working stocks, along with the sequences and binding sites of primers and the expected size of the amplicons, are documented as part of the written laboratory protocol. Polymorphisms or translocation breakpoints that affect primer binding should be noted in terms of the expected frequency in the population or in the number of successful amplifications.

For hybridization procedures, labeled probe solutions are treated as working stock and verified by parallel analysis with old lots. FISH probes are validated and verified according to recommended procedures. The quality of new microarray lots is verified by the manufacturer or by hybridizing labeled nucleotides that bind to all probes on a representative array from the lot. RNA probes are maintained under RNase-free conditions to protect their integrity.

Descriptive information on probes is documented in the laboratory. This information includes the type of probe (genomic, cDNA, oligonucleotide, plasmid, or riboprobe) and the species of origin of the probe sequence. The sequence of the probe, a GenBank number or other identification of the target sequence or gene region recognized by the probe, and a restriction enzyme map of that region are also important information. Any known polymorphisms, sites resistant to endonuclease digestion, and cross-hybridizing bands should be noted. Recombination frequencies and map positions must be documented for linkage procedures. For inherited disease tests, the chromosomal location of the target gene and known mutant alleles and their frequencies in various ethnic groups might be cited in published reports. Labeling methods and standards for adequacy of hybridization are included in the test procedure manual.

In multiplex reactions, primer and/or probe competition for substrates may affect the results. Multiple fluorochrome signals in fluorescence assays may also cross into each other's detection ranges. For gel or capillary sizing, products of multiplex reactions should be reasonably different so that banding patterns do not complicate interpretation. For example, multiplex STR analysis by PCR includes 13 sets of primers that produce 13 amplicons labeled with three different fluorochromes. The range of sizes of each amplified STR locus is designed not to overlap others labeled with the same fluorochrome. The instrument that detects the fluorescence is also calibrated to subtract any overlap of detection of one fluorochrome with another.

Reagent Categories

The Medical Device Amendment to the Federal Food, Drug and Cosmetic Act (1938) authorized the FDA to regulate medical devices, including laboratory tests. Subsequent acts and amendments have further defined the control, risk, safety, and performance of medical devices. A growing list of categories defines the degree of endorsement of the performance of tests and test components by the FDA.

Analyte-specific reagents (ASRs) are probes, primers, antibodies, and other test components that detect a specific target, such as a cell surface protein or DNA mutation. ASRs comprise the active part of "homebrew" tests and are usually purchased from an outside manufacturer. ASRs are classified as I, II, or III. Most ASRs used in the molecular laboratory are class I. Several molecular tests are available as ASRs in infectious disease, tissue typing, and other areas of molecular diagnostics. Approved molecular methods include tests that utilize FISH, hybrid capture, PCR, and microarray technologies. Class II and III ASRs include those used by blood banks to screen for infectious diseases and those used in the diagnosis of certain contagious diseases, such as tuberculosis. Class I ASRs are not subject to special controls by the FDA. The test performance of class I ASRs is established during the validation of individual LDTs.

In the diagnostic laboratory, **in vitro diagnostic (IVD)** devices or reagent sets are intended for use in the diagnosis of disease or other conditions. IVD reagent sets include products used to collect, prepare, and examine specimens collected from patients. They should be used strictly according to the manufacturer's protocol.

IVD devices are classified according to performance risk and requirements for surveillance and verification, from class I (lowest risk) to class III (highest risk). Regardless of risk level, the performance (accuracy, precision, analytic sensitivity, reportable range, reference range) of IVD tests must be verified by the laboratory in which they will be used. If an IVD protocol is modified, the procedure must be validated to show equivalent or better performance compared with the original assay. A list of currently available FDA-approved tests is located on the Association for Molecular Pathology website (https://www.amp.org/).

The FDA would approve the analytical validity only of a proposed category of **in vitro analytical test (IVAT)** reagents, with no claims of clinical utility. The clinical laboratory would be responsible for any clinical claims. This category is intended to accommodate the use of promising high-complexity technologies facing long clearance processes; **research use only (RUO)** and **investigational use only (IUO)** reagents are not intended for diagnostic use. RUO reagents are not intended for use on patient samples, whereas IUO reagent can be used on patient samples with proper institutional review and informed consent, for example, in clinical trials. RUO and IUO reagents are used to gather data that may result in the advancement of a product's status. These types of reagents are used for testing standard analytes, such as viral load.

Chemical Safety

Volatile and flammable reagents used in the molecular laboratory, such as xylene, methanol, ethanol, and isopropanol, are stored in properly vented and explosion-proof cabinets or refrigeration units (Fig. 15.16). The National Fire Protection Association (NFPA) has developed a series of warning labels for universal use on all chemical containers (Fig. 15.17).[20] Secondary or reinforced containers are required for transport and handling of dangerous chemicals, such as concentrated acids and phenol.

Radioactive chemicals are used in some molecular methods (Table 15.5). Although methods involving radiation are increasingly being replaced by nonradioactive alternatives, some laboratory procedures still use these agents. The Nuclear Regulatory Commission requires that laboratory personnel working with radioactive reagents maintain a radiation safety manual providing procedures for the safe handling of radioactive substances in both routine and emergency situations. The Occupational Safety and Health Administration (OSHA) has also developed regulations regarding ionizing and nonionizing radiation.

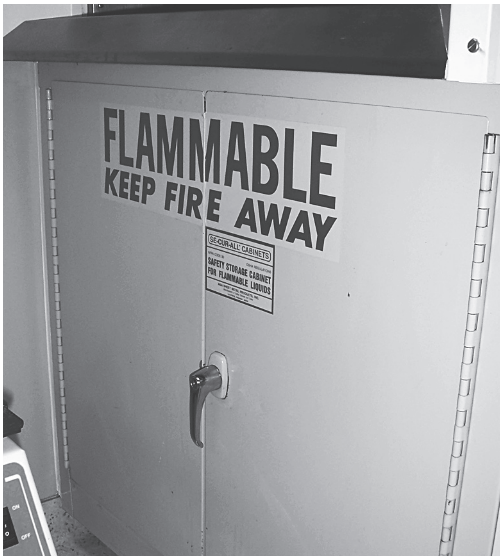


FIGURE 15.16 Flammable and explosive materials are stored in designated protective cabinets or explosion-proof refrigerators.

Radioactive reagents and methods are performed in designated areas. Working surfaces are protected with absorbent paper, drip trays, or other protective containers. Potentially volatile radioactive materials are handled under a fume hood. Radioactive waste is discarded in appropriate containers, separate from normal trash, according to regulations. Some isotopes with short half-lives may be stored over approximately seven half-lives, checked for residual emissions, and then discarded with regular waste. Containers and equipment used in these areas should be labeled with "Caution Radioactive Material" signs (Fig. 15.18). Signs should be posted in the rooms where radioactive materials are used. OSHA has specifications for accident prevention signs and tags for radiation and other occupational hazards.

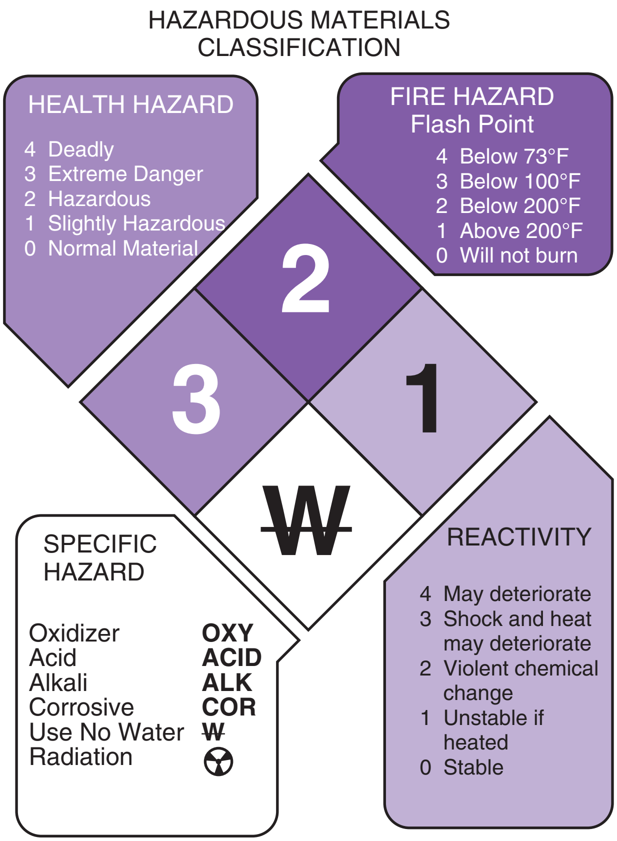

HAZARDOUS MATERIALS CLASSIFICATION
HEALTH HAZARD
4 Deadly
3 Extreme Danger
2 Hazardous
1 Slightly Hazardous
0 Normal Material
FIRE HAZARD
Flash Point
4 Below 73°F
3 Below 100°F
2 Below 200°F
1 Above 200°F
0 Will not burn
2
3
1
₩
SPECIFIC HAZARD
Oxidizer OXY
Acid ACID
Alkali ALK
Corrosive COR
Use No Water ₩
Radiation ☢
REACTIVITY
4 May deteriorate
3 Shock and heat may deteriorate
2 Violent chemical change
1 Unstable if heated
0 Stable

FIGURE 15.17 NFPA hazard labels have three parts, labeled with numbers 0 to 4, depending on the amount of hazard, from none (0) to severe (4). The fourth section has two categories. OXY indicates a strong oxidizer, which greatly increases the rate of combustion. The ₩ symbol indicates dangerous reactivity with water, which would prohibit the use of water to extinguish a fire in the presence of this chemical. See Color Plate 12.

Laboratory personnel working with radioactive material should receive special training for the safe handling, decontamination, and disposal of radiation. Laboratory instructions for working with radiation should include inspection and monitoring of shipments as required by the U.S. Department of Transportation. Workspaces are decontaminated daily and checked at least monthly by swipe testing or by Geiger counter. Technologists wear gloves, lab coat, and safety glasses when handling radioactive solutions. Radiation badges are worn when handling 1.0 mCi or more. Exposure increases with decreasing distance from the radioactive reagent (see Table 15.5), so exposure at close distance, such as working over open containers, should be avoided. For isotopes such as ^{32}P, acrylic shielding is required for work, storage, and waste areas (Fig. 15.19).

Reagent Storage

Proper reagent storage ensures optimal stability and assay performance. The time and temperature of storage should be documented. Reagents for molecular testing are supplied with recommended storage conditions and expiration dates. Expiration dates can be recorded in laboratory documents or on the reagent containers.

Storage conditions should not be altered. Most enzymes, antibodies, and other proteins are stable at freezer temperatures (–18°C to –25°C); however, multiple freeze–thaw cycles will degrade the proteins. Polymerases, endonucleases, and other enzymes supplied in glycerol are protected from the formation of ice crystals that will degrade the proteins. Reagent supplied frozen in larger volumes (1 to 50 mL) can be aliquoted to avoid freezing and thawing the entire amount multiple times. Because thawing is sometimes done on ice, and even at room temperature, the time required for thawing can slow the testing process. Reagents to be stored at refrigerator temperatures (4ºC to 10ºC) may not be stable at freezer temperatures because they may not survive even a few freeze–thaw cycles. Not all proteins are stable at freezer temperatures. Frost-free freezers are not recommended for reagent and specimen storage because the defrosting cycle can thaw small reagent volumes.

Freeze-dried or anhydrous reagents are stored at room temperature, unless otherwise specified. In some areas, humidity may have to be monitored where dry and dried reagents are stored.

Reagent Labeling

In addition to the time and temperature of storage, reagent labels provide chemical and safety information, including the chemical name, signal words (danger or warning), a pictogram of hazard symbols, the manufacturer/supplier's contact information, and precautionary and hazard statements. The United States adopted the Globally Harmonized System (GHS) of Classification and Labeling of Chemicals in 2012 as an update to the Occupational Health and Safety Act. The system was fully

TABLE 15.5 Examples of Radionuclides Used in Laboratory Methods

Radioisotope	Half-Life*	Radiation (MeV)†	Travel in Air	Critical Organs
^{32}P	14.29 days	β, 1.709	20 feet	Bone, whole body
^{33}P	25.3 days	β, 0.249	20 feet	Bone, whole body
^{3}H	12.35 years	β, 0.019	0.65 inches	Body water
^{14}C	5,730 years	β, 0.156	10 inches	Whole body, fat
^{35}S	87.39 days	β, 0.167	11 inches	Whole body, testes
^{125}I	60.14 days	X, γ, 0.035	3 feet	Thyroid

*Time for half of the radiation emission to dissipate.
†Million electron volts.

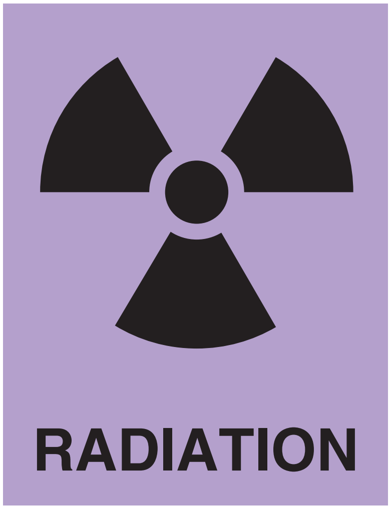


FIGURE 15.18 Rooms, cabinets, and equipment containing radioactive chemicals are identified with radiation safety labels. See Color Plate 13.

implemented in June 2016. This system requires updating of chemical labeling and standardization of Material Safety Data Sheets (MSDSs), now designated Safety Data Sheets (SDSs). SDSs are available to all laboratory personnel in hard-copy or electronic form.

Reagents aliquoted or diluted into secondary containers must also be labeled with the reagent name, hazard symbols, and, if diluted, concentration and diluent, if other than water. Primary and secondary labels should indicate expiration dates.

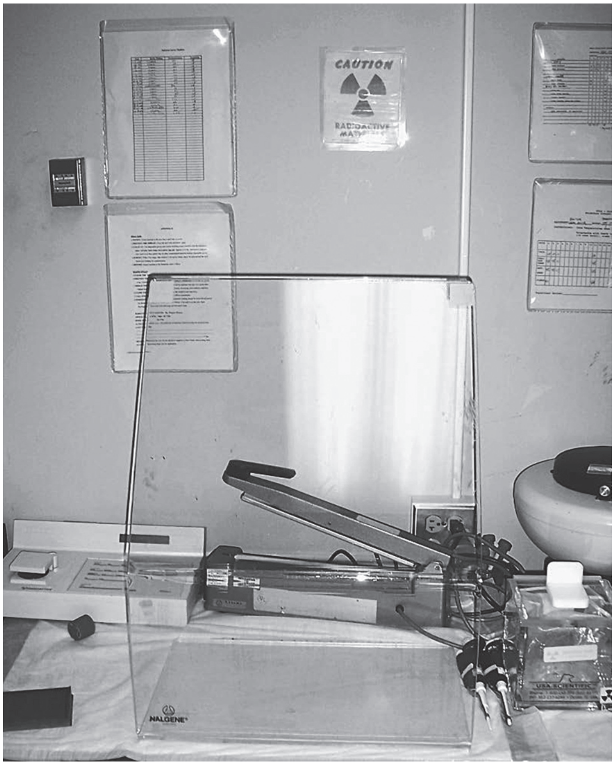


FIGURE 15.19 Acrylic shielding is required for working with gamma emitters, such as ^{32}P and ^{33}P.

A chemical hygiene plan for the laboratory should include responsibilities for lab personnel, chemical handling instructions, PPE usage, exposure monitoring requirements, and employee training. Signage posted at the laboratory entrance should indicate the presence of carcinogens such as ethidium bromide or radioactive materials, irritants such as xylene, flammable substances such as **reagent alcohol,** and other hazards.[21]

PROFICIENCY TESTING

Proficiency testing refers to the analysis of external specimens from a reference source supplied to independent laboratories. Proficiency testing is performed to assess the skills (competency) of laboratory personnel performing molecular assays as well as the performance of the assay itself. The availability of comprehensive test specimens in the rapidly expanding area of molecular diagnostics is sometimes problematic. The CAP supplies specimens for molecular oncology, engraftment, and microsatellite instability testing, among others (http://www.cap.org). Several analytes, however, are not available, especially for tests that are offered in a small number of laboratories. If proficiency specimens are not commercially available, laboratories can exchange blinded split specimens; alternatively, blinded specimens measured or documented by independent means, such as chart review, can be tested within the laboratory. Increasing numbers of reference standards that can be used for proficiency testing are commercially available.[22,23] If at all possible, inter-laboratory testing is preferable.

Proficiency testing is performed at least twice a year, with the proficiency samples tested within routine patient runs. Handling, analysis, review, and reporting of proficiency test results are included in a written laboratory proficiency testing procedure. The specific procedures should be defined and documented in the laboratory. Errors or incorrect responses for proficiency specimens are documented, along with the corrective action taken, if necessary.

DOCUMENTATION OF TEST RESULTS

Test results in the form of electropherograms, gel images, and autoradiograms should be of a sufficiently high quality that results are unequivocal. This includes clear bands or peaks without high background, cross-hybridization, distortions, and other artifacts. Controls should also be clear and consistent and reflect the expected size or level. Molecular-weight ladders on gels, autoradiograms, or electropherograms should cover the expected range of band or peak sizes produced from the specimen. For example, if primers used to detect a t(14;18) translocation test by PCR yield amplicons expected to range from 150 to 500 bp, the molecular-weight ladder used must range from less than 150 bp to more than 500 bp.

A record of the assay conditions and reagent lot numbers is kept with patient results. Identification of the technologist performing the assay may also be included. Documentation of the quality and quantity of the isolated DNA or RNA is also required, especially if designated amounts of nucleic acids are used for an assay. Quantity and quality are documented in the form of spectrophotometry or fluorometry data or quantity or gel photographs of high-molecular-weight DNA or ribosomal RNA quality. The quality of RNA analyzed by northern blot or RT-PCR may be assessed by monitoring a housekeeping gene, ribosomal RNA expression, or other calibrator. New lots (or new shipments from the same lot) should be tested with controls before use.

If DNA is cut with restriction enzymes for Southern blot or PCR-restriction fragment length polymorphism, complete cutting by the restriction enzyme is verified on control targets and documented by photography of the cut DNA on the gel after electrophoresis. DNA digested with DNase for array analysis is also documented in this way to confirm proper fragment sizes. If specimen nucleic acid is labeled for hybridization arrays, labeling efficiency is assessed by measurement of the specific activity (signal per ng nucleic acid). For Southern blots, patient identification, gel lane (well) number, and probe target and type are also documented. It is also recommended that the test documentation should include pre-hybridization and hybridization conditions and probe and hybridization buffer lot numbers.

In situ results, such as FISH, are correlated with histological findings (stained sections) of tissue morphology. This is required when the molecular target detection is significant in specific cells, for example, with p53 detection in tumor cells. Documentation includes images of at least one normal cell with at least two abnormal cell results. These images are cross-referenced or retained together with photographs, films, and **autoradiographs** generated from additional testing of the same specimen. All these records are labeled with patient identification, sample numbers, run identifiers, and the date of assay.

All of the raw data are retained with the final report and clinical interpretation of the test results. Careful documentation is important because molecular diagnostic results may differ from results from other laboratories or from the clinical diagnosis. Such discrepancies occur most often with amplification methods because of their high sensitivity. If a molecular result is questioned, investigation of the discrepancy includes a review of the raw data. The results of this investigation, along with any corrective action taken, are noted in the laboratory records.

Gene Nomenclature

Reporting of results from genetic testing requires unequivocal identification of the gene being tested. Gene names are sometimes not consistent. There can be multiple names referring to the same gene. For example, the official HUGO Gene Nomenclature Committee (HGNC) gene name *ERBB2* is also called *HER2, HER-2,* and *EGFR2*. Even the official name can be changed by the HGNC. Ongoing efforts to implement the universal usage of standard official gene names and symbols by molecular laboratory professionals are documented at http://www.genenames.org. Meanwhile, laboratory reports may refer to "also known as" names for the same gene, for example, *ERBB2 (HER2).* Gene name aliases can be found at the National Center for Biological Information (NCBI) Gene Database (http://www.ncbi.nlm.nih.gov/gene/). It is most important to use the gene name or symbol recognized by the laboratory's medical community.

Gene Sequencing Results

Direct sequencing is increasingly used in clinical applications to detect gene mutations or to type microorganisms. Sequence data must be of adequate quality with acceptably low baseline, especially if heterozygous target states are to be detected. Each nucleotide peak or band should be unequivocal. Sequencing should be performed on both complementary strands of the template to confirm sequenced mutation or type. Repeated sequencing across the same area, or **resequencing,** for known sequence changes is sometimes performed only on one template strand; however, sequencing of both strands is best.

The criteria for the acceptance of sequencing data include correct assignment of the nucleotide sequence in a defined region surrounding the critical area, not including the amplification primer-binding sites. Furthermore, a specified level of band or peak quality (intensity or fluorescence levels, respectively) with a reasonably low background is assigned. Defined limits of fluorescence ratios are set to identify true heterozygous base positions. Ideally, a heterozygous position will have equal fluorescence contribution from the two genotypes, and the peak height will be approximately half that of a homozygous genotype at that position. Results are expressed in the standard nomenclature for DNA or protein sequences (see Chapter 9).

The utility of sequence data requires published normal or type-specific sequences. In the case of gene mutations, electronic or published databases of known mutations and polymorphisms are available for frequently tested genes. These records, especially Internet databases, are updated regularly. Newly discovered mutations are classified according to the type of mutation; the laboratory director or consultant uses published guidelines to determine if the mutation is clinically significant.[24] For example, a silent mutation will not affect protein function, whereas a frameshift mutation will.

REPORTING RESULTS

Test results are reported in terms that are unambiguous to readers who may not be familiar with molecular methods or terminology. The test report must clearly convey the method or manufactured kit used, the locus, the mutation or organism tested, the analytical interpretation of the raw data, and the clinical interpretation of the analytical result. This interpretation includes the penetrance of mutations, that is, the probability of having a mutation but not getting the associated disease.

The likelihood of false-positive or false-negative results is also included in a report. Mutation detection is not guaranteed, especially in large genes with hundreds of possible mutations that may or may not effectively compromise gene function. The mutation detection rate for this type of gene and the residual risk of undetected mutations are therefore included in the test report. Negative results from tests for specific point or chromosomal mutations are reported in terms of the sensitivity of the test (e.g., less than 0.01% chance of mutation) or, alternatively, accompanied with the sensitivity levels of the test. For parentage reports, the combined paternity index, the probability of paternity as a percentage, the prior probability of paternity used in calculations, and the population used for comparison are reported.

The laboratory director, pathologist, or other clinical expert reviews the analytical interpretation, determines the clinical interpretation, and verifies the final results with an actual or electronic signature on the test report. An internal laboratory summary sheet or database is often useful for compiling pertinent information. Test results should not be released before they are reviewed by the director. Molecular diagnostic tests, in particular, may have technical complexities that influence the meaning of the test result. These results are best communicated with the clinical significance of the laboratory findings.

When class I ASRs are used in an analytical method, the following disclaimer is included in the test report: "This test was developed, and its performance characteristics determined by (laboratory name). It has not been cleared or approved by the U.S. Food and Drug Administration." Additional language is also recommended by the CAP: "The FDA does not require this test to go through premarket FDA review. This test is used for clinical purposes. It should not be regarded as investigational or for research. This laboratory is certified under the Clinical Laboratory Improvement Amendments (CLIA) as qualified to perform high complexity clinical laboratory testing."[25]

The disclaimer is not required for tests using reagents that are sold together with other materials or an instrument as a reagent set nor for reagents sold with instructions for use.

Maintaining the confidentiality of molecular test results is essential. All results, and particularly molecular genetic results, may affect insurability, employment, or other family members. Results are released only to the ordering physician or other authorized personnel such as genetic counselors or nurse coordinators. Technologists should refer requests for patient data to supervisors. Data sent by facsimile must be accompanied by a disclaimer such as: "The documents accompanying this telecopy transmission contain confidential information belonging to the sender that is legally privileged. This information is intended only for the use of the individual or entity named above. The authorized recipient of this information is prohibited from disclosing this information to any other party and is required to destroy the information after its stated need has been fulfilled. If you are not the intended recipient, you are hereby notified that any disclosure, copying, distributing, or action taken in reliance on the contents of these documents is strictly prohibited. If you have received this telecopy in error, please notify the sender immediately to arrange for return or destruction of these documents."

Test results are not released to employers, insurers, or other family members without the patient's expressed consent. Any data discussed in a public forum are presented such that no patient or pedigree is identifiable by the patient or the general audience. Written consent from the patient may be required under some circumstances. Each institution will have a department that oversees the lawful use of confidential information.

Technologists working in the area of molecular pathology will encounter tests in which the final details are determined empirically. As a result, a test procedure may differ from one laboratory to another. Even after the test procedure is established, troubleshooting is sometimes required as the procedure is used on a routine basis. Some reactions that work well for short-term research may prove to be less consistent and reproducible than is required in the clinical laboratory setting.

Biotechnology is quickly developing standard reagent sets and instrumentation for the most popular tests, but these also differ from one supplier to another. Furthermore, due to market demands, test reagent kits may be modified or discontinued. If replacement reagents are available, they may not be identical to those previously used. Ongoing tests then must be optimized. This can be a concern where turnaround times are critical.

It then becomes the responsibility of the technologist to perform and monitor tests on a regular basis to maintain the consistency and accuracy of results. The technologist who understands the biochemistry and molecular biology of these tests will be better able to respond to these problems. In addition, with the quickened evolution of the sciences, a knowledgeable technologist can better recognize significant discoveries that offer the potential for test improvement.

STUDY QUESTIONS

What actions should be taken in the following situations?

1. An unlabeled collection tube with a requisition for a factor V Leiden test is received in the laboratory.
2. After PCR, the amplification control has failed to yield a product.
3. An isolated DNA sample is to be stored for at least 6 months.
4. A bone marrow specimen arrives at the end of a shift and will not be processed for the *Bcl2* translocation until the next day.
5. The temperature of a refrigerator set at 8°C (±2°C) reads 14°C.
6. A PCR test for the *BCR/ABL* translocation was negative for the patient sample and for the sensitivity control.
7. A fragile X test result has been properly reviewed and reported.
8. A bottle of reagent alcohol with a 3 in the red diamond on its label is to be stored.
9. The expiration date on a reagent has passed.
10. Test results are to be faxed to the ordering physician.

References

1. Li S, Chen H, Song J, Lee C, Geng Q. Avoiding heparin inhibition in circulating MicroRNAs amplification. *International Journal of Cardiology* 2016;207:92–93.
2. Hecker K, Taylor PD, Gjerde DT. Mutation detection by denaturing DNA chromatography using fluorescently labeled polymerase chain reaction products. *Analytical Biochemistry* 1999;272: 156–164.
3. Lam N, Rainer TH, Chiu RWK, Lo YMD. EDTA is a better anticoagulant than heparin or citrate for delayed blood processing for plasma DNA analysis. *Clinical Chemistry* 2004;50:256–257.
4. Tanner M, Berk LS, Felten DL, Blidy AD, Bit SL, Ruff DW. Substantial changes in gene expression level due to the storage temperature and storage duration of human whole blood. *Clinical and Laboratory Haematology* 2002;24:337–341.
5. Chai V, Vassilakos A, Lee Y, Wright JA, Young AH. Optimization of the PAXgene^{TM} blood RNA extraction system for gene expression analysis of clinical samples. *Journal of Clinical Laboratory Analysis* 2005;19:182–188.
6. Duale N, Lipkin WI, Briese T, Aarem J, Rønningen KS, Aas KK, Magnus P, Harbak K, Susser E, Brunborg G. Long-term storage of blood RNA collected in RNA stabilizing Tempus tubes in a large biobank—evaluation of RNA quality and stability. *BMC Research Notes* 2014;7:633.
7. Lou J, Mirsadraei L, Sanchez DE, Wilson RW, Shabihkhani M, Lucey GM, Wei B, Singer EJ, Mareninov S, Yong WH. A review of room temperature storage of biospecimen tissue and nucleic acids for anatomic pathology laboratories and biorepositories. *Clinical Biochemistry* 2014;47:267–273.
8. Takeo S, Ishii A, Segawa T, Takagi Y, Kobayashi Y, Itou T. Establishing conditions for the storage and elution of rabies virus RNA using FNA® cards. *Journal of Veterinary Medicine* 2015;77:461–465.
9. Burd E. Validation of laboratory-developed molecular assays for infectious diseases. *Clinical Microbiological Reviews* 2010;23: 550–576.
10. Pyne M, Konnick EQ, Phansalkar A, Hillyard DR. Evaluation of the Abbott investigational use only RealTime hepatitis C virus (HCV) assay and comparison to the Roche TaqMan HCV analyte-specific reagent assay. *Journal of Clinical Microbiology* 2009;47: 2872–2878.
11. Goedel S, Rullkoetter M, Weisshaar S, Mietag C, Leying H, Boehl F. Hepatitis B virus (HBV) genotype determination by the COBAS AmpliPrep/COBAS TaqMan HBV Test, v2.0 in serum and plasma matrices. *Journal of Clinical Microbiology* 2009;45: 232–236.
12. Jennings L, Van Deerlin VM, Gulley ML; College of American Pathologists Molecular Pathology Resource Committee. Recommended principles and practices for validating clinical molecular pathology tests. *Archives of Pathology & Laboratory Medicine* 2009;133:743–755.
13. De Vore K, Agrwal Y, Alspach TD, Budd JR, Durazo-Arvizu RA, Eckfeldt JHJ, Goodwain K, Halim A, Long T, Miller, WG, Pamapori N, Thomas J, Vaks JE, Vesper HW. Evaluation of commutability of processed samples; EP14-A3. *CLSI Approved Guideline* 2014;34:1–41.
14. Gabert J, Beillard E, van der Velden VH, Bi W, Grimwade D, Pallisgaard N, Barbany G, Cazzaniga G, Cayuela JM, Cavé H, Pane F, Aerts JL, De Micheli D, Thirion X, Pradel V, González M, Viehmann S, Malec M, Saglio G, van Dongen JJ. Standardization and quality control studies of "real-time" quantitative reverse transcriptase polymerase chain reaction of fusion gene transcripts for residual disease detection in leukemia—a Europe Against Cancer program. *Leukemia* 2003;17:2318–2357.
15. Coskun A. Modified Levey-Jennings charts for calculated laboratory tests. *Clinical Chemistry and Laboratory Medicine* 2006;44:387–389.

16. Chang W, McLean IP. Cusum: a tool for early feedback about performance? *BMC Medical Research Methodology* 2006;6:1–5.

17. Carroll T, Pinnick HA, Carroll WE. Probability and the Westgard rules. *Annals of Clinical Laboratory Science* 2003;33:113–114.

18. Ayling P, Hill RP, Jassam N, Kallner A, Khatami Z. A practical tool for monitoring the performance of measuring systems in a laboratory network: report of an ACB Working Group. *Annals of Clinical Biochemistry* 2017;54(6):702–706.

19. Lasky F. *Evaluation of matrix effects; approved guideline*. 2nd ed. Wayne, PA: Clinical and Laboratory Standards Institute; 2005.

20. National Fire Protection Association. Hazardous materials code. NFPA 400; 2016.

21. Scungio D. Safety first with chemical handling and labeling. *Clinical Laboratory News* 2017;43:4–5.

22. Parisien M, Wang X, Pan T. Diversity of human tRNA genes from the 1000-Genomes project. *RNA Biology* 2013;10:1853–1867.

23. Sarawitz S, George H, Miller WG, Tholen DW, Valenstein P. *Clinical and Laboratory Standards Institute document GP29-A2: assessment of laboratory tests when proficiency testing is not available; approved guideline*. 2nd ed. Wayne, PA: Clinical and Laboratory Standards Institute; 2008.

24. den Dunnen JT, Dalgleish R, Maglott DR, Hart RK, Greenblatt MS, McGowan-Jordan J, Roux A, Smith T, Antonarakis SE, Taschner PE. HGVS recommendations for the description of sequence variants. *Human Mutation* 2017;37(6):564–569.

25. American College of Pathologists. *Molecular Pathology Checklist*. Northfield, IL: American College of Pathologists; 2016.

Appendix A

Study Question Answers

Chapter 1 Study Question Answers

DNA Structure and Function

1. What is the function of DNA in the cell?

   Answer: DNA is a storage system for genetic information.

2. Compare the structure of the nitrogen bases. How do purines and pyrimidines differ?

   Answer: Purines have a double ring; pyrimidines have a single ring.

3. Write the complementary sequence to the following:

   5′AGGTCACGTCTAGCTAGCTAGA3′

   Answer: 3′TCCAGTGCAGATCGATCGATCT5′

4. Which of the ribose carbons participate in the phosphodiester bond?

   Answer: The 5′ carbon connects to a hydroxyl group on the 3′ ribose carbon of the previous nucleotide.

5. Which of the ribose carbons carries the nitrogen base?

   Answer: The 1′ carbon carries the nitrogen base.

6. Why does DNA polymerase require primase activity?

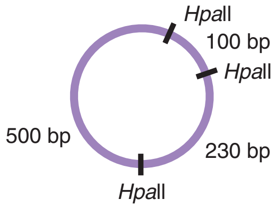
   Answer: A 3′ hydroxyl group from an existing nucleotide must be present to form the phosphodiester bond.

Restriction Enzyme Analysis

1. A plasmid was digested with the enzyme *Hpa*II. On agarose gel electrophoresis, you observe three bands, 100, 230, and 500 bp.

   Answer:

   a. How many *Hpa*II sites are present in this plasmid? **There are three *Hpa*II sites.**
   b. What are the distances between each site? **100 bp, 230 bp, 500 bp**
   c. What is the size of the plasmid? **The plasmid is 830 bp in size.**
   d. Draw a picture of the plasmid with the *Hpa*II sites.



A second cut of the plasmid with *Bam*H1 yields two pieces, 80 and *x* bp.

   e. How many *Bam*H1 sites are in the plasmid? **There are two *Bam*H1 sites.**
   f. What is *x* in base pairs (bp)? **830 bp – 80 bp = 750 bp**

2. How would you determine where the *Bam*H1 sites are in relation to the *Hpa*II sites?

   Answer: Cut the plasmid with both enzymes at the same time.

3. The plasmid has one *Eco*R1 site into which you want to clone a blunt-ended fragment. What type of enzyme could turn an *Eco*R1 sticky end into a blunt end?

   Answer: A 5′ to 3′ single-strand exonuclease will remove the single-stranded overhangs.

Recombination and DNA Transfer

1. Describe how DNA moves from cell to cell by (a) conjugation, (b) transduction, and (c) transformation.

   Answer:

   a. **Cells share DNA by direct cell-to-cell contact.**
   b. **Viral or bacteriophage vectors carry DNA from cell to cell.**
   c. **Fragmented or plasmid DNA enters cells from the surrounding environment.**

2. Which of the three interactions in question 1 would be prevented by a barrier between two mating strains that stops bacterial cells but not smaller particles?

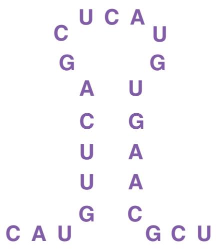
   Answer: Conjugation would be prevented because there would be no cell-to-cell contact. Smaller viral particles, plasmids, and fragments of DNA would pass through the membrane.

3. After meiosis, gametes produced from diploid organisms are **haploid** (haploid/diploid).

RNA Secondary Structure

1. Draw the secondary structure of the following RNA. The complementary sequences (inverted repeat) are underlined.

   Answer: 5′CAUGUUCAGCUCAUGUGAAC GCU3′



2. Underline *two* inverted repeats in the following RNA.

   Answer: 5′CUGAACUUCAGUCAA GCAAAGAGUUUGCACUG3′

RNA Processing

1. Name three processing steps undergone by messenger RNA.

   Answer: mRNA is capped during transcription and polyadenylated as part of termination. The transcript is further processed by splicing.

2. What is the function of polyadenylate polymerase?

   Answer: This enzyme adds adenosine nucleotides to the 3′ end of mRNA.

3. What is unusual about the phosphodiester bond formed between mRNA and its 5′ cap?

   Answer: The phosphodiester bond involved in capping is a 5′–5′ bond, rather than the more usual 3′–5′ bond.

4. The parts of the hnRNA that are translated into protein (open reading frames) are the **exons**.

5. The intervening sequences that interrupt protein-coding sequences in hnRNA are **introns**.

Proteins and the Genetic Code

1. Indicate whether the following peptides are hydrophilic or hydrophobic.

   Answer:

   a. MLWILAA **hydrophobic**
   b. VAIKVLIL **hydrophobic**
   c. CSKEGCPN **hydrophilic**
   d. SSIQKNET **hydrophilic**
   e. YAQKFQGRT **hydrophilic**
   f. AAPLIWWA **hydrophobic**
   g. SLKSSTGGQ **hydrophilic**

2. Is the following peptide positively or negatively charged at neutral pH?

   GWWMNKCHAGHLNGVYYQGGTY

Answer: The peptide is positively charged with basic side chains pI > pH 7.

3. Consider an RNA template made from a 2:1 mixture of C:A. What would be the three amino acids *most* frequently incorporated into the protein?

   Answer: Proline (CCA), histidine (CAC), and threonine would be frequently incorporated.

4. What is the peptide sequence encoded in AUAUAUAUAUAUAUA . . . ?

   Answer: The protein would alternate isoleucine (AUA) and tyrosine (UAU).

5. Write the *anticodons* 5′ to 3′ of the following amino acids:

   Answer:

   a. L **UAA, CAA, AAG, GAG, UAG, CAG**
   b. T **AGU, GGU, UGU, CGU**
   c. M **CAU**
   d. H **AUG, GUG**
   e. R **ACG, GCG, UCG, CCG**
   f. I **AAU, GAU**

6. A protein contains the sequence LGEKKWCLRVNPKGLDESKDYLSLKSKYLLL. What is the likely function of this protein? (Note: See Advanced Concepts box.)

   Answer: This protein has leucine (L) residues at sequential seventh positions, forming a leucine zipper found in transcription factors (DNA-binding proteins).

7. A histone-like protein contains the sequence PKKGSKKAVTKVQKKDGKKRKRSRK. What characteristic of this sequence makes it likely to associate with DNA?

   Answer: The positive charge on this protein, facilitating association with the negatively charged DNA, makes it likely to associate with DNA.

8. A procedure for digestion of DNA with a restriction enzyme includes a final incubation step of 5 minutes at 95°C. What is the likely purpose of this final step?

   Answer: The 95°C incubation will inactivate the protein, preventing its activity in subsequent steps of the assay.

9. What is a ribozyme?

   Answer: A ribozyme is an RNA molecule that can metabolize other molecules like an enzyme.

10. Name the nonprotein prosthetic groups for the following conjugated proteins:

glycoprotein	**sugars**
lipoprotein	**lipids**
metalloprotein	**metal atoms**

Chapter 2 Study Questions

Gene Expression

1. Proteins that bind to DNA to control gene expression are called **trans (or transcription)** factors.

2. The binding sites on DNA for proteins that control gene expression are __**cis**__ factors.

3. How might a single mRNA produce more than one protein product?

   Answer: A single RNA may be alternatively spliced to produce more than one protein product.

4. The type of transcription producing RNA that is continually required and relatively abundant in the cell is called __**constitutive**__ transcription.

5. A set of structural genes transcribed together on one polycistronic mRNA is called an __**operon**__.

The Lac Operon

Using the depiction of the lac operon in Figure 2.4, indicate whether gene expression (transcription) would be on or off under the following conditions: (P = promoter; O = operator; R = repressor).

Answer:

- a. P+ O+ R+, no inducer present—**OFF**
- b. P+ O+ R+, inducer present—**ON**
- c. P– O+ R+, no inducer present—**OFF**
- d. P– O+ R+, inducer present—**OFF**
- e. P+ O– R+, no inducer present—**ON**
- f. P+ O– R+, inducer present—**ON**
- g. P+ O+ R–, no inducer present—**ON**
- h. P+ O+ R–, inducer present—**ON**
- i. P– O– R+, no inducer present—**OFF**
- j. P– O– R+, inducer present—**OFF**
- k. P– O+ R–, no inducer present—**OFF**
- l. P– O+ R–, inducer present—**OFF**
- m. P+ O– R–, no inducer present—**ON**
- n. P+ O– R–, inducer present—**ON**
- o. P– O– R–, no inducer present—**OFF**
- p. P– O– R–, inducer present—**OFF**

Epigenetics

1. Indicate whether the following events would increase or decrease expression of a gene:

   Answer:

   - a. **Methylation of cytosine bases 5′ to the gene would decrease expression.**
   - b. **Histone acetylation close to the gene would increase expression.**
   - c. **siRNAs complementary to the gene transcript will decrease expression.**

2. How does the complementarity of siRNA to its target mRNA differ from that of miRNA?

   Answer: The complementarity of siRNA to its target is exact. The complementarity of miRNA to its target is imperfect.

3. What is imprinting in DNA?

   Answer: DNA imprinting is the marking of selected regions of DNA, usually by methylation.

4. What sequence structures in DNA, usually found 5′ to structural genes, are frequent sites of DNA methylation?

   Answer: CpG islands (more than the expected frequency of CG dinucleotides within a given length of DNA), often found 5′ to structural genes, are targets for methylation.

5. What is the RISC?

   Answer: The RNA-induced silencing complex mediates the specific interaction between siRNA or miRNA and the target RNA.

6. Name four functions of lncRNAs.

   Answer: lncRNA recruit modifiers of chromatin, act as decoys that titrate away DNA-binding proteins such as transcription factors, act as scaffolds to bring two or more proteins into a complex, or bind to DNA in an enhancer-like control of gene expression from distal cis elements.

Chapter 3 Study Questions

DNA Quantity/Quality

1. Calculate the DNA concentration in μg/mL from the following information:

   Answer:

   - a. Absorbance reading at 260 nm from a 1:100 dilution = 0.307

     0.307 × 50 μg/mL = 15.35 μg/mL
     15.35 μg/mL × 100 = 1,535 μg/mL

   - b. Absorbance reading at 260 nm from a 1:50 dilution = 0.307

     0.307 × 50 μg/mL = 15.35 μg/mL
     15.35 μg/mL × 50 = 767.5 μg/mL

   - c. Absorbance reading at 260 nm from a 1:100 dilution = 0.172

     0.172 × 50 μg/mL = 8.60 μg/mL
     8.60 μg/mL × 100 = 860 μg/mL

   - d. Absorbance reading at 260 nm from a 1:100 dilution = 0.088

     0.088 × 50 μg/mL = 4.40 μg/mL
     4.40 μg/mL × 100 = 440 μg/mL

2. If the volume of the DNA solutions in question 1 was 0.5 mL, calculate the yield for (a)–(d).

   Answer:

   a. **1,535 μg/mL × 0.50 mL = 767.5 μg**
   b. **767.5 μg/mL × 0.50 mL = 383.8 μg**
   c. **860 μg/mL × 0.50 mL = 430 μg**
   d. **440 μg/mL × 0.50 mL = 220 μg**

3. Three DNA preparations have the following A_{260} and A_{280} readings:
   For each sample, based on the A_{260}/A_{280} ratio, is each preparation suitable for further use? If not, what is contaminating the DNA?

   Sample OD_{260} OD_{280}
   (i) 1 0.419 0.230
   0.419/0.230 = 1.82.
   This DNA is suitable for use.

   (ii) 2 0.258 0.225
   0.258/0.225 = 1.15.
   This DNA is not suitable for use due to protein contamination.

   (iii) 3 0.398 0.174
   0.398/0.174 = 2.29.
   This DNA may be suitable for use if RNA does not interfere with the subsequent assay.

4. After agarose gel electrophoresis, a 0.5-microgram aliquot of DNA isolated from a bacterial culture produced only a faint smear at the bottom of the gel lane. Is this an acceptable DNA sample?

   Answer: This amount of DNA should produce a bright band/smear near the top of the gel lane. This DNA is probably degraded and is therefore unacceptable.

5. Compare and contrast the measurement of DNA concentration by spectrophotometry with analysis by fluorometry with regard to staining requirements and accuracy.

   Answer: Spectrophotometry requires no DNA staining. Fluorometry requires staining of DNA to generate a fluorescent signal. Fluorometry may be more accurate than spectrophotometry because double-stranded DNA must be intact to stain and generate a signal, whereas single nucleotides will absorb light in spectrophotometry.

RNA Quantity/Quality

1. Calculate the RNA concentration in μg/mL from the following information:

   Answer:

   a. Absorbance reading at 260 nm from a 1:100 dilution = 0.307

      0.307 × 40 μg/mL = 12.28 μg/mL
      12.28 μg/mL × 100 = 1,228 μg/mL

   b. Absorbance reading at 260 nm from a 1:50 dilution = 0.307

      0.307 × 40 μg/mL = 12.28 μg/mL
      12.28 μg/mL × 50 = 614 μg/mL

   c. Absorbance reading at 260 nm from a 1:100 dilution = 0.172

      0.172 × 40 μg/mL = 6.88 μg/mL
      6.88 μg/mL × 100 = 688 μg/mL

   d. Absorbance reading at 260 nm from a 1:100 dilution = 0.088

      0.088 × 40 μg/mL = 3.52 μg/mL
      3.52 μg/mL × 100 = 352 μg/mL

2. If the volume of the RNA solutions in question 1 was 0.5 mL, calculate the yield for (a)–(d).

   Answer:

   a. **1,228 μg/mL × 0.50 mL = 614 μg**
   b. **614 μg/mL × 0.50 mL = 307 μg**
   c. **688 μg/mL × 0.50 mL = 344 μg**
   d. **3.52 μg/mL × 0.50 mL = 1.76 μg**

3. An RNA preparation has the following absorbance readings:

   $A_{260} = 0.208$
   $A_{280} = 0.096$

   Is this RNA preparation satisfactory for use? **The A_{260}/A_{280} ratio is 0.208/0.096 = 2.17. This RNA preparation is satisfactory for use.**

4. A blood sample was held at room temperature for 5 days before being processed for RNA isolation. Will this sample likely yield optimal RNA?

   Answer: This sample will not yield optimal RNA due to degradation and changes in gene expression at room temperature.

5. Name three factors that will affect the yield of RNA from a paraffin-embedded tissue sample.

   Answer: Isolation of RNA from fixed tissue is affected by (1) the type of fixative used, (2) the age/length of storage of the tissue, and (3) the preliminary handling of the original specimen. Other factors such as time of fixation, type of specimen, and specimen transport conditions will also affect RNA quality and yield.

Chapter 4 Study Questions

1. You wish to perform an electrophoretic resolution of your restriction enzyme–digested DNA. The sizes of the expected fragments range from 100 to 500 bp. You discover two agarose gels polymerizing on the bench. One is 0.5% agarose; the other is 2% agarose. Which one might you use to resolve your fragments?

   Answer: The 2% gel is best for this range of fragment sizes.

2. After completion of the electrophoresis of DNA fragments along with the proper molecular-weight standard on an agarose gel, suppose the outcomes in (a) and (b) were observed. What might be the explanations for each outcome?

   Answer:

   a. The gel is blank (no bands, no molecular-weight standard).
   Because the molecular-weight standard is not visible, something is wrong with the general electrophoresis process. Most likely, staining with a DNA-specific dye was omitted.
   b. Only the molecular-weight standard is visible.
   The presence of the molecular-weight standard indicates that the electrophoresis and staining were performed properly. In this case, the DNA fragments were not loaded onto the gel or the method used to produce the fragments was not successful.

3. How does PFGE separate larger fragments more efficiently than standard electrophoresis?

   Answer: PFGE forces large fragments through the gel matrix by repeatedly changing the direction of the current, thus aligning and realigning the particles with the gel spaces.

4. A 6% solution of 19:1 acrylamide:bis-acrylamide is mixed, de-aerated, and poured between glass plates for gel formation. After an hour, the solution is still liquid. What might be one explanation for the gel not polymerizing?

   Answer: The nucleating agent and/or polymerization catalyst were not added to the gel solution.

5. A gel separation of RNA yields aberrantly migrating bands and smears. Suggest two possible explanations for this observation.

   Answer: RNA degradation will yield the aberrantly migrating bands and smears. Improper electrophoresis conditions (buffer, denaturation agent, or gel) will also affect band migration.

6. Why does DNA not resolve well in solution (without a gel matrix)?

   Answer: Particle movement in solution is based on the charge/mass ratio. As the mass of DNA increases, slowing migration, the negative charge increases, counteracting the effect of mass.

7. Why is SYBR green less toxic than EtBr?

   Answer: SYBR green is a minor groove–binding dye. It does not disrupt the nucleotide sequence of DNA. Ethidium bromide (EtBr) is an intercalating agent that slides in between the

nucleotide bases in the DNA, which can cause changes in the nucleotide sequence (mutations) in the DNA.

8. What are the general components of loading buffer used for introducing DNA samples to submarine gels?

   Answer: Gel loading buffers contain a density agent (Ficoll, glycerol, sucrose) to facilitate placing the liquid sample in the gel well underneath the buffer surface and a tracking dye to follow the migration of the sample during electrophoresis.

9. Name two dyes that are used to monitor the migration of nucleic acid during electrophoresis.

   Answer: Bromphenol blue and xylene cyanol green are two of several tracking dyes.

10. When a DNA fragment is resolved by slab gel electrophoresis, a single sharp band is obtained. What would the equivalent observation be if this fragment had been fluorescently labeled and resolved by capillary electrophoresis?

    Answer: Comparable capillary electrophoresis results will be a single peak on an electropherogram.

Chapter 5 Study Questions

1. Calculate the melting temperature of the following DNA fragments using the sequences only:

   a. AGTCTGGGACGGCGCGGCAATCGCA
   TCAGACCCTGCCGCGCCGTTAGCGT

   Answer: (8 × 2°C) + (17 × 4°C) = 84°C

   b. TCAAAAATCGAATATTTGCTTATCTA
   AGTTTTTAGCTTATAAACGAATAGAT

   Answer: (20 × 2°C) + (6 × 4°C) = 64°C

   c. AGCTAAGCATCGAATTGGCCATCGTGTG
   TCGATTCGTAGCTTAACCGGTAGCACAC

   Answer: (14 × 2°C) + (14 × 4°C) = 84°C

   d. CATCGCGATCTGCAATTACGACGATAA
   GTAGCGCTAGACGTTAATGCTGCTATT

   Answer: (15 × 2°C) + (12 × 4°C) = 78°C

2. What is the purpose of denaturation of a double-stranded target DNA after electrophoresis and prior to transfer in a Southern blot?

   Answer: Double-stranded target DNA is denatured to allow hybridization of the probe. Also, single-stranded DNA binds to the nitrocellulose fibers in the blotting membrane.

3. Name two ways to permanently bind nucleic acid to nitrocellulose following transfer.

   Answer: To permanently bind the nucleic acid, bake the filter at 80°C for 30 minutes. Alternatively, nucleic acid will be cross-linked to nitrocellulose upon exposure to ultraviolet light.

4. If a probe for a Southern blot is dissolved in a hybridization buffer that contains 50% formamide, is the stringency of hybridization higher or lower than if there was no formamide?

   Answer: The stringency is higher because formamide facilitates denaturation of double-stranded DNA.

5. If a high concentration of NaCl was added to a hybridization solution, how would the stringency be affected?

   Answer: The stringency would be decreased because high salt concentrations exclude nucleic acids, forcing them close together and promoting hybridization.

6. Does an increase in temperature from 65°C to 75°C during hybridization raise or lower the stringency?

   Answer: Increasing the temperature raises the stringency. Heat promotes separation of the hydrogen bonds holding the two single strands of double-stranded DNA together.

7. At the end of the Southern blot procedure, what would the autoradiogram show if the stringency was too high?

Answer: If the stringency was too high, faint or no bands will be present on the autoradiogram.

8. A Northern blot is performed on an RNA transcript with the sequence GUAGGUATGUAUUUGGGCGCGAACGCAAAA. The probe sequence is GUAGGUATGUAUUUGGGCGCG. Will this probe hybridize to the target transcript?

   Answer: This probe will not bind to the target transcript. The sequences are identical and not complementary. Because RNA does not have a complementary partner, the probe must be complementary to the transcript sequence.

9. In an array CGH experiment, three test samples were hybridized to three microarray chips. Each chip was spotted with eight gene probes (Genes *A*–*H*). The following table shows the results of this assay expressed as the ratio of test DNA to reference DNA. Are any of the eight genes consistently deleted or amplified in the test samples? If so, which ones?

Gene	Sample 1	Sample 2	Sample 3
A	1.06	0.99	1.01
B	0.45	0.55	0.43
C	1.01	1.05	1.06
D	0.98	1.00	0.97
E	1.55	1.47	1.62
F	0.98	1.06	1.01
G	1.00	0.99	0.99
H	1.08	1.09	0.90

   Answer: Gene *B* is consistently deleted (test/reference < 1.0), and gene *E* is consistently amplified (test/reference > 1.0).

10. What are two differences between CGH arrays and expression arrays?

    Answer: CGH arrays are performed on DNA, whereas expression arrays use RNA.

    CGH arrays detect deletions and amplifications of chromosomal regions, whereas expression arrays detect changes in the RNA levels of specific genes.

Chapter 6 Study Questions

1. A master mix of all components (except template) necessary for PCR contains what basic ingredients?

   Answer: An example master mix (without template) will contain the following:
   primers
   dATP, dCTP, dGTP, dTTP
   KCl or other monovalent cation
   Tris buffer
   $MgCl_2$
   polymerase

2. The final concentration of *Taq* polymerase is to be 0.01 units/µL in a 50-µL PCR. If the enzyme is supplied at 5 units/µL, how much enzyme would you add to the reaction?

   a. 1 µL
   b. 1 µL of a 1:10 dilution of *Taq*
   c. 5 µL of a 1:10 dilution of *Taq*
   d. 2 µL

   Answer: Using ratio and proportion to achieve the concentration of 0.01 units/µL, the enzyme should be 0.01/5.00 = x/50. Because x = 0.1, 1 µL of a 1:10 dilution of *Taq* (b) is the correct answer.

3. Primer dimers result from

   a. high primer concentrations.
   b. low primer concentrations.
   c. high GC content in the primer sequences.
   d. 3′ complementarity in the primer sequences.

   Answer: Primer dimers result from 3′ complementarity in the primer sequences (d). Without this “self-priming” of primers, concentration and GC content will not generate PCR products.

4. Which control is run to detect contamination?

   a. Negative control
   b. Positive control
   c. Molecular-weight marker
   d. **Reagent blank**

   Answer: The reaction mix without template, which is the reagent blank (d), is included to detect contamination. Negative and positive controls monitor the performance and accuracy of the reaction. The molecular-weight marker allows assessment of the PCR product size and monitors the gel performance.

5. Nonspecific extra PCR products can result from

   a. **mispriming.**
   b. high annealing temperatures.
   c. high agarose gel concentrations.
   d. omission of $MgCl_2$ from the PCR.

   Answer: Nonspecific products result from mispriming (a). High annealing temperatures and omission of magnesium from the reaction would lessen the production of amplicons. The agarose gel concentration is not related to the PCR reaction.

6. Using which of the following is an appropriate way to avoid PCR contamination?

   a. High-fidelity polymerase
   b. Hot-start PCR
   c. **A separate area for PCR setup**
   d. Fewer PCR cycles

   Answer: A separate pre-PCR setup area (c) with separate equipment and reagents is an appropriate way to prevent contamination. High-fidelity polymerase will copy the template with minimal PCR sequence artifacts but cannot distinguish contamination from the desired template. Hot-start PCR can inhibit mispriming but not contamination. Using fewer PCR cycles will lower the amount of final product but does not prevent misprimed amplification.

7. How many copies of a target are made after 30 cycles of PCR?

   a. 2×30
   b. **2^{30}**
   c. 30^2
   d. 30/2

   Answer: Assuming 100% efficiency, there should be 2^{30} (b) copies of the target after 30 PCR cycles. This reflects doubling of the amplifiable template with each cycle.

8. What are the three steps of a standard PCR cycle?

   Answer: A standard three-step PCR cycle includes a denaturation step where the double-stranded template becomes single-stranded for binding of primers in the second or annealing step of the cycle. Extension or addition of nucleotides to the 3′ ends of the primers is the third step, resulting in two double-stranded copies of the original template.

9. Which of the following is a method for purifying a PCR product?

   a. Treating with uracil N glycosylase
   b. Adding divalent cations
   c. **Putting the reaction mix through a spin column**
   d. Adding DEPC

   Answer: Post-PCR steps of protocols may be affected by residual materials and buffer components from the PCR reaction. Putting the reaction mix through a spin column (c) will collect the PCR product, which can be eluted in a purer solution. N glycosylase will digest DNA containing uracil, which can purify non–uracil-containing DNA, but this mix of DNAs is unlikely to occur in the PCR product. Divalent cations are necessary for polymerase enzyme activity to produce the PCR product but not purify it. DEPC is an agent that prevents RNase from digesting RNA. Although RNA is not a component product of a PCR reaction, if present, it would not be removed by DEPC.

10. In contrast to standard PCR, real-time PCR is
    a. **quantitative.**
    b. qualitative.
    c. labor-intensive.
    d. sensitive.

    Answer: Real-time PCR (qPCR) is quantitative (a). It is not labor-intensive compared with standard PCR because no post-PCR electrophoresis is required to see the product. qPCR can be very sensitive but not necessarily more so than standard PCR.

11. In real-time PCR, fluorescence is *not* generated by which of the following?
    a. FRET probes
    b. TaqMan probes
    c. SYBR green
    d. ***Tth* polymerase**

    Answer: *Tth* polymerase (d) is a DNA replication enzyme and does not produce a signal. Probes and fluorescent dyes are used for product detection in qPCR.

12. Prepare a table that compares PCR, LCR, bDNA, TMA, Qβ replicase, and hybrid capture with regard to the type of amplification, target nucleic acid, type of amplicon, and major enzyme(s) for each.

Amp Method	Target (DNA or RNA)	Amplified Product
PCR, LAMP, HDA, RMA	DNA, RNA	Target (DNA)
TMA (NASBA)	RNA, DNA	Target (RNA)
Qβ replicase	DNA	Probe (DNA, RNA)
LCR	DNA	Probe (DNA)
bDNA	DNA, RNA	Signal
Hybrid Capture	DNA, RNA	Signal

13. Examine the following sequence. (The complementary strand is not shown.) You are devising a test to detect a mutation at the underlined position.

    Answer: There are multiple primers that can satisfy these requirements. One primer pair example is given.

    5′TATTTAGTTA TGGCCTATAC ACTATTTGTG
    AGCAAAGGTG ATCGTTTTCT GTTTGAGATT
    TTTATCTCTT GATTCTTCAA AAGCATTCTG
    AGAAGGTGAG ATAAGCCCTG AGTCTCAGCT
    ACCTAAGAAA AACCTGGATG TCACTGGCCA
    CTGAGGAGC T **TTGTT TCAAC CAAGTCATGT**> GCATTTCCAC

    GTCAACAGAA TTGTTTATTG TGACAGTT**A**T
    ATCTGTTGTC CCTTTGACCT TGTTTCTTGA

    AGGTTTCCTC GTCCCTGGG C AATTCCGCAT
    TTAATTCAT_G GTATTCAGGA

    <G TTAAGGCGTAAATTAAGTA

    TTACATGCAT GTTTGGTTA AACCCATGAGA
    TTCATTCAGT TAAAAATCCA
    GATGGCGAAT3′

Design one set of primers (forward and reverse) to generate an amplicon containing the underlined base.

The primers should be 20 bases long.
The amplicon must be 100 to 150 bp in size.
The primers must have similar melting temperatures (T_m), +/– 2°C.
The primers should have no homology in the last three 3′ bases.

a. Write the primer sequences 5′→3′ as you would if you were to order them from the DNA synthesis facility.

Answer: There are multiple answers to this question. One example:

Forward primer: 5′TTGTT TCAACCAAGTCATGT_3′

Reverse primer: 5′ATGAATTAAATGCGGAATTG3′

b. Write the T_m for each primer that you have designed.

Answer: Forward primer: (13 × 2°C) + (7 × 4°C) = 54°C

Reverse primer: (14 × 2°C) + (6 × 4°C) = 52°C

(Ideally, forward and reverse primer pairs should have similar annealing temperatures ±2°C.)

14. How does nested PCR differ from multiplex PCR?

    Answer: Nested PCR is done in two rounds; that is, the PCR product of round one is used as the template for round two. Multiplex PCR uses more than one primer pair in a single round of PCR.

15. What replaces heat denaturation in strand displacement amplification?

    Answer: Enzymatic nicking of the double-stranded probe product produces a 3′ end for copying the uncut strand while displacing its complement.

Chapter 7 Study Questions

1. What chromosomal location is indicated by 15q21.1?

   Answer: This location is on the long arm of chromosome 15, region 2, band 1, subband 1.

2. During interphase FISH analysis for the t(9;22) translocation, one nucleus was observed with two normal signals (one red for chromosome 22 and one green for chromosome 9) and one composite red/green signal. Five hundred other nuclei were normal. What is one explanation for this observation?

   Answer: The composite signal indicates the presence of a translocation between chromosomes 9 and 22; however, because only 1 of 500 nuclei showed the composite signal, the more likely explanation is that this is an artifactual result, possibly due to the overlap of chromosomes on the slide.

3. Is 47;XYY a normal karyotype?

   Answer: No. This karyotype is aneuploid with polysomy Y (one extra Y chromosome).

4. Write the numerical and structural chromosomal abnormalities represented by the following genotypes:

47,XY, +18	**trisomy 18**
46,XY, del(16) p(14)	**deletion in the short arm of chromosome 16 at region 1, band 4**
iso(Xq)	**isochromosome formed by centromeric joining of two long arms of the X chromosome**
46,XX del(22) q(11.2)	**deletion in the long arm of chromosome 22 at region 1, band 1, subband 2**
45,X	**loss of one X or the Y chromosome**

5. A chromosome with a centromere located such that one arm of the chromosome is longer than the other arm is called

   a. metacentric.
   b. paracentric.
   c. telocentric.
   d. **submetacentric.**

   Answer: A chromosome with a long and short arm is submetacentric (d). In a metacentric chromosome, the centromere is near the middle so that the arms are of approximately equal length. One arm is very short in telocentric chromosomes where the centromere is close to one end. *Paracentric* refers to an inversion within one arm of a chromosome, not involving the centromere.

6. A small portion from the end of chromosome 2 has been found on the end of chromosome 15, replacing the end of chromosome 15, which has moved to the end of chromosome 2. This mutation is called a(n)

   a. **reciprocal translocation.**
   b. inversion.
   c. deletion.
   d. robertsonian translocation.

   Answer: The exchange of portions of chromosomes with no loss of genetic material

is a reciprocal translocation (a). An inversion reverses the orientation of DNA, only involving one chromosome. Deletions are a loss of material, from 1 bp to millions of bp from chromosomes. In robertsonian translocations, chromosomes break at their centromeres, and the long arms fuse to form a single chromosome with one centromere.

7. Phytohemagglutinin is added to a cell culture when preparing cells for karyotyping. The purpose of the phytohemagglutinin treatment is to

   a. arrest the cell in metaphase.
   b. spread out the chromosomes.
   c. fix the chromosomes on the slide.
   d. **stimulate mitosis in the cells.**

   Answer: Cells must enter metaphase for karyotyping. Mitosis is stimulated in culture using phytohemagglutinin (d). Once the cells go into the cell cycle, Colcemid is used to arrest the cells in metaphase. Chromosomes are distributed from the nucleus by lysis with hypotonic buffer. Chromosomes are fixed on the slide with methanol.

8. A centromeric probe is used to visualize chromosome 21. Three fluorescent signals are observed in the cell nuclei when stained with this probe. These results would be interpreted as consistent with

   a. a normal karyotype.
   b. **Down syndrome.**
   c. Klinefelter syndrome.
   d. technical error.

   Answer: Three centromeric signals instead of two from chromosome 21 is a finding consistent with Down syndrome (trisomy 21; b). The presence of two centromeric signals for each chromosome is normal. Klinefelter syndrome is indicated by an extra X chromosome in men (47;XXY). Technical error can result in aberrant signals, such as chromosome overlap; however, these artifactual signals are usually rare.

9. Cells were harvested from a patient's blood, cultured to obtain chromosomes in metaphase, fixed onto a slide, treated with trypsin, and then stained with Giemsa. The resulting banding pattern is called

   a. **G banding.**
   b. Q banding.
   c. R banding.
   d. C banding.

   Answer: Giemsa staining produces G banding (a). When chromosomes are stained with the fluorescent dyes quinacrine and quinacrine mustard, the resulting fluorescence pattern visualized after staining is Q banding. Treatment of chromosomes with acridine orange dye will produce a pattern opposite to the G banding pattern called R banding. Alkali treatment of chromosomes results in centromere staining, or C banding.

10. A FISH test with a centromere 13 probe is ordered for a suspected case of Patau syndrome (trisomy 13). How many signals per nucleus will result if the test is positive for Patau syndrome?

    Answer: The FISH results will reveal three signals per nucleus with a probe to chromosome 13.

11. What would be the results if a centromere 13 probe was used on a case of Edward syndrome (trisomy 18)?

    Answer: The FISH results will reveal two signals per nucleus with a probe to chromosome 13. A probe to centromere 18 would yield three signals per nucleus.

12. Angelman syndrome is caused by a microdeletion in chromosome 15. Which method, karyotyping or metaphase FISH, is better for accurate detection of this abnormality? Why?

    Answer: Metaphase FISH is preferred over karyotyping for the detection of microdeletions. The lower resolution of karyotyping makes the detection of small deletions difficult.

13. The results of a CGH analysis of Cy3 (green)-labeled test DNA with Cy5 (red)-labeled reference

DNA on a normal chromosome spread revealed a bright red signal along the short arm of chromosome 3. How is this interpreted?

a. **3p deletion**
b. 3q deletion
c. 3p amplification
d. 3q amplification

Answer: The red signal from the reference chromosome region indicates a loss or deletion in the test chromosome at 3p (a). Amplification would yield a green signal at this location. Events at 3p would not affect the signal at 3q.

14. A break-apart probe is used to detect a translocation. The results of FISH analysis show two signals in 70% of the nuclei counted and three signals in 30% of the nuclei. Is there a translocation present?

Answer: A translocation is present, indicated by the nuclei in which three signals appear. The probe spans the translocation breakpoint, producing two separate signals when a translocation occurs. The third signal is the intact homologous chromosome.

15. What FISH technique is most useful for the detection of multiple complex genomic mutations?

Answer: Spectral karyotyping labels each chromosome with a different fluorescent color so that multiple complex genomic mutations are more clearly identified.

Chapter 8 Study Questions

1. What characteristic of the genetic code facilitates identification of open reading frames in DNA sequences?

Answer: Out-of-frame or chance consecutive codons tend to be short, often ending in a stop codon.

2. Compare and contrast EIA with Western blots for the detection of protein targets.

Answer: The EIA method involves liquid handling and is more easily automated and analyzed than the Western blot.

3. On a size-exclusion column, large molecules will elute___**before**_____ (before/after) small molecules.

4. MALDI methods separate ions by

a. molecular volume.
b. mass.
c. charge.
d. **mass and charge.**

Answer: MALDI methods separate ions by mass and charge (d), regardless of volume. Molecules to be assessed are ionized before they are attracted to travel through a magnetic field. Low-mass ions and more highly charged ions move faster through the drift space than ions with higher mass and lower charge. Thus, the time of ion flight differs according to the mass-to-charge ratio (m/z).

5. What is a heteroduplex?

Answer: A heteroduplex is one double-stranded DNA molecule with one or more noncomplementary bases.

6. Exon 4 of the *HFE* gene from a patient suspected to have hereditary hemochromatosis was amplified by PCR. The G to A mutation, frequently found in hemochromatosis, creates an *Rsa*1 site in exon 4. When the PCR products are digested with *Rsa*1, what results (how many bands) would you expect to see if the patient has the mutation if no other *Rsa*1 sites are naturally present in the PCR product?

Answer: Digestion with *Rsa*I would produce two bands if the patient has the mutation and if no other *Rsa*I sites are naturally present in the PCR product.

7. Which of the following methods identifies the presence of a mutation but not the mutant sequence?

a. SSP-PCR
b. **SSCP**
c. PCR-RFLP
d. NGS

Answer: SSCP screens for mutations by changes in conformers so that the presence

of a mutation is detected but not the mutant sequence (b). SSP-PCR relies on primers designed to bind the mutant base, and RFLP utilizes restriction enzymes with recognition sites containing the potentially mutated base. In both cases, the mutant sequence is known. NGS allows detection of many variant bases in a sequence context.

8. What is the effect on the protein when a codon sequence is changed from TCT to TCC?

   Answer: There would be no effect on the codon sequence because TCT and TCC both code for the same amino acid, serine. Codon usage may, however, affect translation efficiency.

9. A reference sequence, ATGCCCTCTGGC, is mutated in malignant cells. The following mutations in this sequence have been described. Express these mutations using the accepted gene nomenclature (A = nucleotide position 1).

ATGCGCTCTGGC	**5C>G**
ATGCCCTC - -GC	**9_10del or 9_10delTG**
ATAGCCTCTGGC	**3_4delGCinsAG or 3_4delinsAG**
ATGTCTCCCGGC	**4_9inv6**

10. A reference peptide, MPSGCWR, is subject to inherited alterations. The following peptide sequences have been reported. Express these mutations using the accepted nomenclature (M = amino acid position 1).

MPSTGCWR	**s3_g4inst or s3_g4ins1**
MPSGX	**c5x**
MPSGCWLVTGX	**r7inslvtgx or r7ins5 or r7* or r7lfsx5**
MPSGR	**c5rdel2**
MPSGCWGCWR	**w6_r7insgcw or w6_r7ins3**

Chapter 9 Study Questions

1. Read 5′ to 3′ the first 15 bases of the sequence in the gel on the right in Figure 9.7 (p. 229).

   Answer: The gel sequence is read from the bottom of the gel to the top: 5′ATCGTCCCTAAGTCA3′

2. After an automated dye primer sequencing run, the electropherogram displays consecutive peaks of the following colors:
   red, red, black, green, green, blue, black, red, green, black, blue, blue, blue
   If the computer software displays the fluors from ddATP as green, ddCTP as blue, ddGTP as black, and ddTTP as red, what is the sequence of the region given?

   Answer: Based on the peak colors, the sequence is 5′TTGAACGTAGCCC3′.

3. A dideoxy sequencing electropherogram displays bright (high, wide) peaks of fluorescence, obliterating some of the sequencing peaks. What is the most likely cause of this observation? How might it be corrected?

   Answer: The likely cause is the presence of unincorporated labeled dideoxynucleotides or dye blobs. Cleaning the DNA ladder with spin columns, ethanol precipitation, or bead binding will correct this problem.

4. In a manual sequencing reaction, the sequencing ladder on the polyacrylamide gel is very bright and readable at the bottom of the gel, but the larger (slower-migrating) fragments higher up are very faint. What is the most likely cause of this observation? How might it be corrected?

   Answer: The loss of longer products is caused by an overly high dideoxynucleotide/deoxynucleotide ratio. The problem can be corrected by lowering the concentrations of dideoxynucleotides in the sequencing reaction.

5. In an analysis of the *TP53* gene for mutations, the following sequences were produced. For each sequence, write the expected sequence of the opposite strand that would confirm the presence of the mutations detected.

   Answer:

   5′TATCTGTTCACTTGTGCCCT3′ (Normal)
   5′TATCTGTTCATTTGTGCCCT3′ (Homozygous substitution)
   5′AGGGCACAAATGAACAGATA3′
   5′TATCTGT(T/G)CACTTGTGCCCT3′ (Heterozygous substitution)

5′AGGGCACAAGTG(C/A)ACAGATA3′
5′TATCTGTT(C/A)(A/C)(C/T)T(T/G)(G/T)(T/G)(G/C)CC(C/T)(T/… 3′) (Heterozygous deletion)
5′… /A)(A/G)GG(G/C)(C/A)(A/C)(C/A)A(A/G)(G/T)(T/G)AACAGATA3′

6. A sequence, TTGCTGCGCTAAA, may be methylated at one or more of the cytosine residues. After bisulfite sequencing, the following results are obtained:

   Bisulfite treated: TTGUTGCGUTAAA
   Write the sequence showing the methylated cytosines as C$^{\mathrm{Me}}$.

   Answer: TTGCTGC$^{\mathrm{Me}}$GCTAAA

7. In a pyrosequencing readout, the graph shows peaks of luminescence corresponding to the addition of the following nucleotides:
   dT peak, dC peak (double height), dT peak, dA peak
   What is the sequence?

   Answer: Based on the peak order and heights, the sequence is TCCTA.

8. Why is it necessary to add adenosine residues in vitro to ribosomal RNA before capture for sequencing?

   Answer: mRNA naturally has a polyA 3′ terminus required for immobilization by polyT hybridization. RNA species without a polyA tail must be treated with polyA polymerase to add the 3′ polyA tail.

9. Which of the following is next-generation sequencing?

   a. Maxam–Gilbert
   b. Tiled microarray
   c. Dideoxynucleotide chain terminator sequencing
   d. **Reversible dye terminator sequencing**

   Answer: Reversible dye terminator (d) is one example of NGS technology. Maxam–Gilbert sequencing and tiled arrays are not used in clinical sequencing applications. Dideoxynucleotide chain termination sequencing (Sanger sequencing) technology is used for lower-throughput sequencing protocols.

10. Which of the following projects would require next-generation sequencing?

    a. Mapping a mutation in the hemochromatosis gene
    b. Sequencing a viral genome
    c. **Characterizing a diverse microbial population**
    d. Typing a single bacterial colony

    Answer: Characterizing a diverse microbial population (c) would require high-throughput (massive parallel) sequencing. Analysis of single genes or small genomes can be done with standard Sanger sequencing or pyrosequencing. Bacterial colony typing can be performed by MALDI-TOF or biochemical methods.

Chapter 10 Study Questions

1. Consider the following STR analysis.

Locus	Child	Mother	AF1	AF2
D3S1358	15/⑮	15	15	15/16
vWA	17/⑱	17	17/18	18
FGA	23/㉔	22/23	20	24
TH01	6/⑩	6/7	6/7	9/10
TPOX	11/⑪	9/11	9/11	10/11
CSF1PO	12/⑫	11/12	11/13	11/12
D5S818	10/⑫	10	11/12	12
D13S317	⑨/10	10/11	10/11	9/11

   a. Circle the child's alleles that are inherited from the father.

   Answer: (See table.)

   b. Which alleged father (AF) is not excluded as the biological parent?

   Answer: Based on the circled alleles, AF2 is not excluded.

2. The following evidence was collected for a criminal investigation.

Locus	Victim	Evidence	Suspect
TPOX	11/12	12, 11/12	11
CSF1PO	10	10, 9	9/10
D13S317	8/10	10, 8/10	9/12
D5S818	9/11	10/11, 9/11	11
TH01	6/10	6/10, 8/10	5/11
FGA	20	20, 20/22	20
vWA	15/17	18, 15/17	15/18
D3S1358	14	15/17, 14	11/12

The suspect is heterozygous at the amelogenin locus.

a. Is the suspect male or female?

Answer: The suspect is male (XY), as indicated by heterozygosity at the amelogenin locus.

b. In the evidence column, circle the alleles belonging to the victim.

Answer: (See table.)

c. Should the suspect be held or released?

Answer: The genotype of the suspect is not consistent with the remaining (un-circled) alleles. The suspect should be released.

3. A child and an alleged father (AF) share alleles with the following paternity index.

Locus	Child	AF	Paternity Index for Shared Allele
D5S818	$\underline{9}$,10	$\underline{9}$	0.853
D8S1179	$\underline{11}$	$\underline{11}$,12	2.718
D16S539	13,$\underline{14}$	10,$\underline{14}$	1.782

a. What is the combined paternity index from these three loci?

Answer: For the shared alleles (underlined), the combined paternity index is:

$$\mathbf{0.853 \times 2.718 \times 1.782 = 4.131}$$

b. With 50% prior odds, what is the probability of paternity from these three loci?

Probability of paternity

$$\mathbf{= (4.131 \times 0.5)/[(4.131 \times 0.5) + 0.5] = 0.80 = 80\%}$$

Answer: At least 8 loci are commonly tested for paternity.

4. Consider the following theoretical allele frequencies for the loci indicated.

Locus	Alleles	Allele Frequency
CSF1PO	14	0.332
D13S317	9,10	0.210, 0.595
TPOX	8,11	0.489, 0.237

a. What is the overall allele frequency for this genotype, using the product rule?

Answer: The allele frequency = 0.332 × 0.332 × 0.21 × 0.595 × 0.489 × 0.237 = 0.001596

b. What is the probability that this DNA found at two sources came from the same person?

Answer: 1/0.001596 = 626.5. It is 626 times more likely that the two DNA samples came from the same person as from two random persons in the population. Legal identity (CODIS) is based on 12 core loci plus amelogenin with recent recommendations for additional loci being added.

5. STR at several loci were screened by capillary electrophoresis and fluorescent detection for informative peaks prior to a bone marrow transplant. The following results were observed.

Locus	Donor Alleles	Recipient Alleles
LPL	7,10	7,9
F13B	8,14	8
FESFPS	10	7
F13A01	5,11	5,11

Which loci are informative?

Answer: Loci LPL and FESFPS are informative. Locus F13B is donor informative. F13A01 is not informative.

6. An engraftment analysis was performed by capillary gel electrophoresis and fluorescence detection. The fluorescence as measured by the instrument under the FESFPS donor peak was 28,118 units, and that under the FESFPS recipient peak was 72,691. What is the percent donor in this specimen?

   Answer:
   $$\%\ \text{Donor} = 28{,}118/(28{,}118 + 72{,}691) = 27.9\%$$

7. The T-cell fraction from the blood sample in question 6 was separated and measured for donor cells. Analysis of the FESFPS locus in the T-cell fraction yielded 15,362 fluorescence units under the donor peak and 97,885 under the recipient peak. What does this result predict with regard to T-cell-mediated events such as graft-versus-host disease or graft-versus-tumor?

   Answer: % Donor = 15,362/(15,362 + 97,885) = 13.6%. This low percentage of T cells predicts fewer T-cell-mediated events.

8. If a child had a Y haplotype including DYS393 allele 12, DYS439 allele 11, DYS445 allele 8, and DYS447 allele 22, what are the predicted Y alleles for these loci of the natural father?

   Answer: The natural father will have the same Y chromosome as the child: DYS393 allele 12, DYS439 allele 11, DYS445 allele 8, and DYS447 allele 22.

9. Which of these would be used for a surname test: Y-STR, Mini-STR, mitochondrial typing, or autosomal STR?

   Answer: Y-STR would be used for a surname or paternal lineage test, going back multiple generations. Mini-STR and autosomal STR are inherited from both parents, making inheritance over many generations complex. Mitochondrial alleles are maternally inherited.

10. An ancient bone fragment was found and said to belong to an ancestor of a famous family. Living members of the family donated DNA for confirmation of the relationship. What type of analysis would likely be used for this test? Why?

    Answer: Mitochondrial DNA typing might be indicated because (1) the small circular, naturally amplified mitochondrial DNA is more likely to be obtained from the old sample, and (2) lineage across several generations can be determined using the maternal inheritance of mitochondrial type.

11. What is a biological exception to positive identification by autosomal STR?

    Answer: Identical twins (and clones) have identical nuclear DNA profiles.

12. A partial STR profile was produced from a highly degraded sample. Alleles matched to a reference sample at five loci. Is this sufficient for a legal identification?

    Answer: Five loci are not sufficient for legal identification. Nonmatching alleles at any of these loci may support exclusion.

13. What is an SNP haplotype? What are tag SNPs?

    Answer: Single-nucleotide polymorphisms (SNPs) that re-inherited together in a block without combination between them comprise an SNP haplotype. SNP haplotypes are identifiable by two or three representative SNPs (tag SNPs) within the haplotype.

14. Which of the following is an example of linkage disequilibrium?
    a. Seventeen members of a population of 1,000 people have a rare disease, and all 17 people have the same haplotype at a particular genetic location on chromosome 3.
    b. Five hundred people from a population of 1,000 people have the same SNP on chromosome 3.

    Answer: Example (a) is linkage disequilibrium, that is, lack of separation of inheritance of the rare disease and the haplotype. Example (b) is a measure of allele frequency.

15. Why are SNPs superior to STR and RFLP for mapping and association studies?

    Answer: SNPs are more numerous in the genome than STRs and RFLPs and, therefore, offer higher resolution for mapping of precise genome locations.

Chapter 11 Study Questions

1. Which of the following genes would be analyzed to determine whether an isolate of *Staphylococcus aureus* is resistant to oxacillin?
    a. ***mecA***
    b. *gyrA*
    c. *inhA*
    d. *vanA*

    Answer: *S. aureus* developed resistance to antibiotics that target its penicillin-binding protein (PBP1) by replacing PBP1 with PBP2a encoded by the *mecA* gene (a). PBP2a found in methicillin-resistant *S. aureus* (MRSA) has a low binding affinity for methicillin. *gyrA* negatively supercoils DNA, favoring replication, transcription, recombination, and repair. *inhA* is an essential enzyme of the mycolic acid biosynthetic pathway in *M. tuberculosis*. The *vanA* operon codes for enzymes that modify the vancomycin-binding site in vancomycin-resistant enterococci (VRE). Transfer of the transposon containing the operon to *S. aureus* produces vancomycin-resistant *S. aureus* (VRSA).

2. Which of the following is a genotypic method used to compare two isolates in an epidemiological investigation?
    a. Biotyping
    b. Serotyping
    c. **Ribotyping**
    d. Bacteriophage typing

    Answer: Of the options offered, only ribotyping (c) is genotypic. Biotyping is a phenotypic biochemical reaction. Serotyping is based on cell surface–antigen phenotype and is used to classify organisms to the subspecies level. Bacteriophage typing is based on the variability between strains of susceptibility to infection by particular phages.

3. For which of the following organisms must caution be exercised when evaluating positive PCR results because the organism can be found as normal flora in some patient populations?
    a. *Neisseria gonorrhoeae*
    b. HIV
    c. *Chlamydophila pneumoniae*
    d. ***Streptococcus pneumoniae***

    Answer: Although PCR is specific for *S. pneumoniae*, the clinical significance of a positive result is not certain (d). A significant proportion of the population (especially children) is colonized with the organism, and PCR cannot discern between colonization and infection. *Neisseria*, HIV, and *Chlamydophila* are rarely found in healthy populations in the absence of risk factors.

4. Which of the following controls are critical for ensuring that amplification is occurring in a patient sample and that the lack of PCR product is not due to the presence of polymerase inhibitors?
    a. Reagent blank
    b. Sensitivity control
    c. Negative control
    d. **Amplification control**

    Answer: The amplification control (d) should always generate a product, even in the

absence of target. The reagent blank monitors for contamination. The sensitivity control demonstrates that the test is operating properly at the lower detection levels. The negative control ensures the specificity of detection of the desired target.

5. A PCR assay was performed to detect *Bordetella pertussis* on sputum obtained from a 14-year-old girl who has had a chronic cough. The results revealed two bands, one consistent with the internal control and the other consistent with the size expected for amplification of the *B. pertussis* target. How should these results be interpreted?
   - a. These are false-positive results for *B. pertussis*.
   - b. **The girl has clinically significant *B. pertussis* infection.**
   - c. *B. pertussis* detection is more likely due to colonization.
   - d. The results are invalid because two bands were present.

   Answer: The girl has clinically significant *B. pertussis* infection (b). Molecular analysis is used almost exclusively for microorganisms such as *N. gonorrhoea, C. trachomatis,* and *B. pertussis*. Because the patient shows symptoms of infection and the target-specific band is of the correct size, this result is unlikely to be a false positive if controls are suitable. Two bands are expected in the positive case, whereas one band (amplification control) would be present for a negative result.

6. Which of the following is an advantage of molecular-based testing?
   - a. **Results stay positive long after successful treatment.**
   - b. Results are available within hours.
   - c. Only viable cells yield positive results.
   - d. Several milliliters of specimen must be submitted for analysis.

   Answer: Molecular testing can be highly sensitive for target nucleic acids, even from dead or residual benign organisms, so that test results may stay positive long after successful treatment (a). Advantages of molecular testing are rapid methods and minimal sample requirements.

7. Which molecular-based typing method has high typing capacity, reproducibility and discriminatory power, moderate ease of performance, and good-to-moderate ease of interpretation?
   - a. Repetitive elements
   - b. **PFGE**
   - c. Plasmid analysis
   - d. PCR-RFLP

   Answer: As shown in Table 11.14, PFGE (b) has the performance characteristics listed. The performance characteristics of repetitive elements, plasmid analysis, and PCR-RFLP are moderate to good; however, PCR-RFLP and repetitive elements are easier to interpret compared with PFGE.

8. A patient has antibodies against HCV and a viral load of 100,000 copies/mL. What is the next test that should be performed on this patient's isolate?
   - a. Ribotyping
   - b. PCR-RFLP
   - c. **Hybrid capture**

   Answer: Hybrid capture (c) can detect a minimum of 4,000 to 5,000 viral genomes. The test also has the capacity for subtyping. Ribotyping and PCR-RFLP might be used for epidemiological studies.

9. A positive result for HPV type 16 indicates
   - a. high risk for antibiotic resistance.
   - b. low risk for cervical cancer.
   - c. **high risk for cervical cancer.**

   Answer: HPV types 16, 18, 31, 33, 35, 39, 45, 51, 56, 58, 59, 68, 73, and 82 have been classified as oncogenic and are found to cause anogenital cancers (c). These types are not associated with specific responses to antibiotic treatment.

10. Which of the following is used to type molds?

    a. Sequence-specific PCR
    b. Microarray
    c. **ITS sequencing**
    d. Flow cytometry

    Answer: Molds are typed by PCR and sequencing of internal transcribed spacer (ITS) regions (c) in 28s RNA. The high resolution of SSP-PCR and array technology is not indicated for identifying molds. Microscopic examination of morphology provides more information than would detection by flow cytometry, which would require a staining antibody for molds.

Chapter 12 Study Questions

1. Which of the following is not a triplet-repeat expansion disorder?

    a. Fragile X syndrome
    b. Huntington disease
    c. **Factor V Leiden**
    d. Congenital central hypoventilation syndrome

    Answer: Factor V Leiden (c) is a point mutation (1691A→G, R506Q) in the *F5* gene and not a disorder. Fragile X syndrome results from expansion of GCC repeats 5′ to the *FMR1* gene. Expansion of a CAG repeat within the *HTT* gene causes Huntington disease. CCHS is associated with a polyalanine (GCN) repeat expansion in the *PHOX2B* gene.

2. A gene was mapped to region 3, band 1, subband 1, of the long arm of chromosome 2. How would you express this location from an idiogram?

    Answer: The designation of this location is 2q31.1.

3. Which of the following can be detected by PCR?

    a. Large mitochondrial deletions
    b. **Full fragile X disorder**
    c. **Mitochondrial point mutations**

    Answer: Whereas full fragile X disorder (b) can be detected with PCR and capillary electrophoresis, and mitochondrial point mutations (c) can be detected by PCR-RFLP or sequencing, large mitochondrial deletions are detected with blot hybridization techniques.

4. A patient was tested for Huntington disease. PCR followed by PAGE revealed 25 CAG units. How should the results be interpreted?

    a. This patient has Huntington disease.
    b. This patient has a 1/25 chance of contracting Huntington disease.
    c. **This patient is normal at the Huntington locus.**
    d. The test is inconclusive.

    Answer: This patient is normal at the HTT locus (c). The frequency of the disorder is 3 to 7 per 100,000 people of European ancestry and less for Asian and African ancestries. In Huntington disease, the CAG repeat expands from 9 to 37 repeats to 38 to 86 repeats.

5. Which of the following methods can detect the factor V Leiden mutation?

    a. PCR-RFLP
    b. SSP-PCR
    c. Invader technology
    d. **All of the above**

    Answer: All of the listed methods (d) are capable of detecting a single-nucleotide change in DNA. PCR-RFLP relies on a restriction enzyme recognition site containing the target nucleotide. In SSP-PCR, the 3′ end of one primer hybridizes to the target nucleotide. Invader technology relies on a probe designed to hybridize to the normal or mutant sequence.

6. The most frequently occurring mutation in the *HFE* gene results in the replacement of cysteine (C) with tyrosine (Y) at position 282. How is this expressed according to the recommended nomenclature?

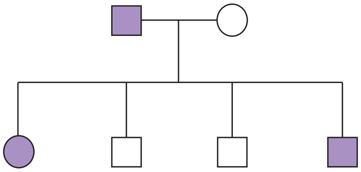
    Answer: The recommended nomenclature is C282Y.

7. MELAS is a disease condition that results from an A to G mutation at position 3243 of the mitochondrial genome. This change creates a single *Apa*I restriction site in a PCR product, including the mutation site. What would you expect from a PCR-RFLP analysis for this mutation in a patient with MELAS?

   a. A single PCR product resistant to digestion with *Apa*I
   b. **A single PCR product that cuts into two fragments upon digestion with *Apa*I**
   c. A single PCR product only if the mutation is present
   d. Two PCR products

   Answer: Amplification of the regions containing the restriction site will yield one amplicon, which will cut into two *Apa*I fragments in the presence of the mutation (b). In the absence of the mutation (and the disease), the PCR product would be resistant to digestion with *Apa*I.

8. A father affected with a single-gene disorder and an unaffected mother have four children (three boys and a girl), two of whom (one boy and the girl) are affected. Draw the pedigree diagram for this family.



   D16S539, an STR, was analyzed in the family. The result showed that the father had the 6,8 alleles, and the mother had the 5,7 alleles. The affected children had the 5,6 and 6,7 alleles, and the unaffected children had the 5,8 and 7,8 alleles.

   a. If D16S539 is located on chromosome 16, where is the gene for this disorder likely to be located?

   Answer: Because all affected individuals have the same D16S539 allele in this limited example, the gene for this disorder is linked to D16S539 on chromosome 16.

   b. To which allele of D16S539 is the gene linked?

   Answer: The gene is linked to allele 6 of D16S539.

   How might one perform a DNA analysis for the presence of the disorder?

   a. **Analyze D16S539 for the 6 allele by PCR.**
   b. Sequence the entire region of the chromosome where D16S539 was located.
   c. Test as many STRs as possible by PCR.
   d. Use a variant-specific test to detect the unknown gene mutation.

   Answer: Because the 6 allele of D16S539 is linked to the disease gene, analysis of the D16S539 locus for the 6 allele (a) would be informative. Sequencing the chromosome around D16S539 and testing many STR are less practical ways to find or detect the disease gene variant. A variant-specific technology would require identification of the variant allele, which is unknown in this case.

9. Exon 4 of the *HFE* gene from a patient suspected of having hereditary hemochromatosis was amplified by PCR. The G to A mutation, frequently found in hemochromatosis, creates an *Rsa*1 site in exon 4. When the PCR products are digested with *Rsa*1, which of the following results would you expect to see if the patient has the mutation?

   a. None of the PCR products will be cut by *Rsa*1.
   b. There will be no PCR product amplified from the patient DNA.
   c. **The patient's PCR product will yield extra bands upon *Rsa*1 digestion.**
   d. The normal control PCR products will yield extra *Rsa*1 bands compared with the patient sample.

   Answer: The patient's DNA should be amplifiable, and the amplicon will yield extra bands upon *Rsa*1 digestion (c). Normal DNA should not be digested at the mutated site.

10. Most people with the C282Y or H63D *HFE* gene mutations develop hemochromatosis symptoms. This is a result of

- a. iron loss.
- b. excessive drinking.
- c. **high penetrance.**
- d. healthy lifestyle.
- e. glycogen accumulation.

Answer: Genetically, if a disease phenotype is frequently present with the DNA variant, then the variant is said to have a high penetrance (c). Other conditions, such as iron loss or glycogen accumulation, are phenotypes that may or may not accompany the hemochromatosis (too much iron). Lifestyle and alcohol intake are behaviors that might affect susceptibility to diseases in general.

11. The majority of disease-associated mutations in the human population are

- a. autosomal dominant.
- b. **autosomal recessive.**
- c. X-linked.
- d. found on the Y chromosome.

Answer: Most individuals carry recessive disease mutations (b) that may result in affected offspring if two disease alleles are inherited (homozygosity). Dominant and sex-linked mutations (which can be hemizygous in males) may result in disease states that negatively affect the reproductive health of a population and survival; thus, they are selected against over time.

12. Bead array technology is most appropriate for which of the following?

- a. **Cystic fibrosis mutation detection**
- b. Chromosomal translocation detection
- c. STR linkage analysis
- d. Restriction fragment length polymorphisms

Answer: Bead arrays are used to detect multiple point mutations such as those found in cystic fibrosis (a). Translocations, STR analysis, and RFLP require size analysis, which is not a capability of bead arrays.

Chapter 13 Study Questions

1. What are the two important checkpoints in the cell division cycle that are crossed when the regulation of the cell division cycle is affected?

Answer: The transition from unreplicated DNA into DNA synthesis (G1 to S) and from replicated DNA to cell division (G2 to M) are regulated checkpoints of the cell division cycle.

2. An EWS-FLI-1 mutation was detected in a solid tumor by RT-PCR. Which of the following does this result support?

- a. Normal tissue
- b. **Ewing sarcoma**
- c. Inherited breast cancer
- d. Microsatellite instability

Answer: The EWS-FLI-1 (t;11;22) translocation is found in Ewing sarcoma (b). A translocation would not likely be inherited. Microsatellite instability results from loss of DNA mismatch repair functions. Driver mutations such as this translocation would transform normal cells and thus would not be found in normal tissue.

3. Mutation detection, even by sequencing, is not definitive with a negative result. Why?

Answer: Mutations may be present outside of the area analyzed. In addition, epigenetic, post-transcriptional, and post-translational events may affect gene function without changing the DNA sequence. Furthermore, abnormalities (gene mutations, epigenetic changes in expression) in factors that interact with the target gene product may affect its function in the absence of mutations in the target gene.

4. A PCR test for the BCL-2 translocation is performed on a patient with suspected follicular lymphoma. The results show a bright band at about 300 bp for this patient. How would you interpret these results?

Answer: A PCR method utilizing one primer on chromosome 18 and one primer on chromosome 14 would only yield a product

if the t(14;18) translocation (found in follicular lymphoma) is present. The presence of the product, therefore, supports the diagnosis of follicular lymphoma. The size of the product, 300 bp, is within an expected amplicon size range, which depends on the primer-binding sites and the translocation breakpoint on chromosome 18.

5. Which of the following misinterpretations would result from PCR contamination?

   a. **False positive for the t(15;17) translocation**
   b. False negative for the t(15;17) translocation
   c. False negative for a gene rearrangement

   Answer: Because PCR contamination is the presence of product, contamination would result in a false-positive interpretation (a). A negative result for a translocation detected by PCR is the absence of a target amplicon. A negative gene rearrangement analysis results in a polyclonal pattern (series of bands or peaks), which usually does not occur from a contaminating product.

6. After amplification of the t(12;21) breakpoint by qRT-PCR, what might be the explanation for each of the following observations? (Assume that positive and amplification controls and a reagent blank control are included in the run.)

   Answer:

   a. All Ct values are reported as undetectable. **There is a problem with the amplification, most likely omission of the detection dye or probe difficulties.**
   b. Sample and control Ct values are very high. **There is a problem with the sample PCR, such as concentration of a critical component required for the PCR reaction.**
   c. There are Ct values detected in the reagent blanks. **The presence of a product in the reagent blank, which should have no template, indicates contamination.**

7. What is observed on a Southern blot for gene rearrangement in the case of a positive result?

   a. No bands
   b. **Germline bands plus rearranged bands**
   c. Smears
   d. Germline bands only

   Answer: Rearranged bands represent the gene rearrangement in the monoclonal population of tumor cells (b). There should be germline bands present unless all the cells in the sample are tumor cells. Smears or no bands on Southern blot are technical difficulties and are not interpretable.

8. Cyclin D1 promotes passage of cells through the G1-to-S checkpoint. What test detects translocation of this gene to chromosome 14?

   a. t(14;18) translocation analysis *(BCL2/IGH)*
   b. t(15;17) translocation analysis *(PML/RARA)*
   c. **t(11;14) translocation analysis *(BCL1/IGH)***
   d. t(8;14) translocation analysis *(MYC/IGH)*

   Answer: Cyclin D1 is encoded by the *BCL1* gene located on chromosome 11. The test, therefore, would have to include analysis of chromosome 11. Translocation to chromosome 14 would be detected as t(11;14) (c).

9. Why is the Southern blot procedure superior to the PCR procedure for detecting clonality in some cases?

   a. Southern blot requires less sample DNA than does PCR.
   b. **The PCR procedure cannot detect certain gene rearrangements that are detectable by Southern blot.**
   c. Southern blot results are easier to interpret than PCR results.
   d. PCR results are not accepted by the College of American Pathologists.

   Answer: The Southern blot can detect all gene rearrangements, regardless of the sequence structure in the immunoglobulin gene regions (b). Detection of rearranged genes by PCR requires that the PCR binding sites are present in the genes involved, which may be lost due to the rearrangement

process and somatic hypermutation. PCR is technically less difficult to perform and requires less sample than Southern blot. Interpretation of PCR results is comparable to interpretation of Southern blot results. Both methods are reviewable by the CAP if properly validated.

10. Interpret the following results from a translocation assay.
    Are the samples positive, negative, or indeterminate?

    Answer:
    Sample 1: This is a negative result.
    Sample 2: This result is indeterminate because the amplification control product is not present. The PCR reaction may not have worked.
    Sample 3: This is a positive result.

11. Which of the following predicts the efficacy of EGFR tyrosine-kinase inhibitors?

    a. Overexpression of EGFR protein
    b. **EGFR-activating mutations**
    c. Patient gender
    d. Stage of disease

    Answer: Evidence suggests that, unlike, for example, the HER2 protein, overexpression of the EGFR protein does not predict the efficacy of EGFR TKI. TKI molecules bind better to the protein product of the *EGFR* gene with activating mutations (b) and are more effective in their presence. The efficacy of TKI is not related to gender or stage of disease.

12. What is the advantage of macrodissection in testing for tumor-specific molecular markers from paraffin-embedded formalin-fixed tissue sections?

    Answer: Macrodissection enriches the representation of tumor cells in the test sample. This increases the sensitivity of the test because normal cells may dilute the tumor-specific markers with unaffected DNA.

13. Why are KRAS- and BRAF-activating mutations almost always exclusive of one another?

    Answer: KRAS and BRAF gene products are on the same signal transduction pathway. Once that pathway is activated through mutations in either gene, there is no selective advantage to generate further activating mutations in the same pathway.

14. What enzyme is responsible for continued sequence changes in the immunoglobulin heavy-chain gene variable region after gene rearrangement has occurred?

    Answer: Activation-induced cytidine deaminase alters C residues so that they mispair, forming heteroduplexes. The mispaired bases are resolved by error-prone repair, which may not insert the correct bases.

15. Why are translocation-based PCR tests more sensitive than IgH, IgL, or T-cell receptor gene-rearrangement tests?

    Answer: Gene rearrangements are not tumor specific, so there will be a background of normal cell gene rearrangements detected by the tests performed with consensus primers. Translocations are tumor specific and not found in normal cells. The absence of background allows for more sensitive detection of tumor cells. Gene rearrangements may be detected with increased sensitivity using patient-specific primers designed to recognize only the tumor gene rearrangement in a given patient, but even this approach is affected by the ongoing evolution of tumor cell populations.

Chapter 14 Study Questions

1. Which of the following is a high-resolution HLA typing result?

    a. B27
    b. A*02:02–02:09
    c. **A*02:12**
    d. A*26:01/A*26:05/A*26:01/A*26:15

    Answer: A*02:12 (c) is a high-resolution typing result indicating the 12th specific allele of the HLA-A *25 family of alleles. B27 is a low-resolution serological typing

result. A*02:02–02:09 and A*26:01/A*26:05/A*26:01/A*26:15 are medium-resolution results.

2. Which of the following is a likely haplotype from parents with A25,Cw10,B27/A23,Cw5,B27 and A17,Cw4,B10/A9,Cw7,B12 haplotypes?

   a. **A25,Cw10,B27**
   b. A25,Cw5,B27
   c. A23,Cw4,B12
   d. A17,Cw4,B27

   Answer: Because alleles in a haplotype do not separate and recombine, only haplotype A25,Cw10,B27 (a) will be inherited from the parents with the listed haplotypes. A25,Cw5,B27 would require two recombination events between A25,Cw10,B27 and A23,Cw5,B27. A17,Cw4,B27 would require a single exchange between A17,Cw4,B10 and A23,Cw5,B27 after fertilization. A23,Cw4,B12 would require exchanges between three of the haplotypes.

3. Upon microscopic examination, over 90% of cells are translucent after a CDC assay. How are these results scored according to the ASHI rules?

   Answer: This observation is interpreted as negative for the presence of the test antibody, with a score of 1. Less than 10% of the cells have taken up the dye. Cytotoxicity (dye uptake) will only occur in those cells that carry antigens to the test antibody.

4. An HLA-A allele is a CTC to CTT (leu → leu) change at the DNA level. How is this allele written?

   a. HLA-A*02
   b. **HLA-A*02:01:01**
   c. HLA-A2
   d. HLA-A*02N

   Answer: The first field of digits after the gene name is the serological type. The next field (after the colon) is the subtype, where nucleotide substitutions that change the amino acid sequence are indicated. A third field of digits, as in HLA-A*02:01:01 (b), designates changes in the DNA sequence that do not change the amino acid sequence. HLA-A*02 is a low-resolution typing. HLA-A2 is not an allele designation. The *N* suffix indicates no expression of the protein.

5. A candidate for kidney transplant has a PRA of 75%. How will this affect eligibility for immediate transplant?

   Answer: A transplant candidate with a %PRA activity of more than 50% is considered to be highly sensitized. Finding a crossmatch-negative donor will be difficult in this case.

6. An SSOP probe recognizes HLA-DRB*03:01–03:04. Another probe recognizes HLA-DRB*03:01/03:04, and a third probe hybridizes to HLA-DRB*03:01–03:03. Test specimen DNA hybridizes to all except the third probe in a reverse dot-blot format. What is the HLA-DRB type of the specimen?

   Answer: The test specimen does not hybridize to HLA-DRB*03:01–03:03, ruling out those types. The other two probes both recognize HLA-DRB*03:04, which is the HLA-DRB type based on these observations.

7. What is the relationship between alleles HLA-A*10 and HLA-A*26(10)?

   Answer: HLA-A*10 is the parent allele of HLA-A*26(10). Number designations of new alleles of a previously defined broad specificity or parent allele follow the parent allele in parentheses.

8. A CDC assay yields an 8 score for sera with the following specificities: A2, A28 and A2, A28, B7, and a 1 score for serum with an A2 specificity. What is the HLA-A type?

   Answer: This is type A28. The high toxicity reading (>6) in the wells containing A2, A28 sera and A2, A28, B7 sera suggest that the cells being tested have surface antigens matching the A28 antibodies.

9. HLA-DRB1*15:01 differs from DRB1*01:01 by a G to C base change. If the sequence surrounding the base change is … GGGTGCGGTTGCTGG AAAGAT … (DRB1*01:01) or … GGGTGCGG

TTCCTGGAAAGAT … (DRB1*15:01), which of the following would be the 3′ end of a sequence-specific primer for detection of DRB1*15:01?

a. … ATCTTTCCAGGAACCC
b. … ATCTTTCCAGCAACCC
c. … ATCTTTCCAGC
d. … ATCTTTCCAGG

Answer: SSP-PCR requires complementarity between the 3′ end of the primer and the template. Primer … ATCTTTCCAGG (d) would detect … GGGTGCGGTTCCTGGAAAGAT … (DRB1*15:01). Primer ATCTTTCCAGC is identical but not complementary to the template sequence. Primers ATCTTTCCAGGAACCC and ATCTTTCCAGCAACCC do not end on the polymorphic base that defines DRB1*15:01.

10. The results of an SSP-PCR reaction are the following: lane 1, one band; lane 2, two bands; lane 3, no bands. If the test includes an amplification control multiplexed with the allele-specific primers, what is the interpretation for each lane?

Answer: Lane 1 is a negative result because the one band is the amplification control. Lane 2 is a positive result for the allele detected in that reaction. Lane 3 is a failed PCR and therefore not informative.

Chapter 15 Study Questions

What actions should be taken in the following situations?

1. An unlabeled collection tube with a requisition for a factor V Leiden test is received in the laboratory.

Answer: Notify the laboratory supervisor and reject the specimen. All primary specimen containers, regardless of accompanying documents, require a label carrying at least two patient-specific identifiers.

2. After PCR, the amplification control has failed to yield a product.

Answer: The results from the PCR cannot be accepted. Check the original nucleic acid preparation by fluorometry or gel electrophoresis. If it is adequate, repeat the amplification. If not, re-isolate the nucleic acid.

3. An isolated DNA sample is to be stored for at least 6 months.

Answer: Store the DNA at –70°C in a tightly sealed tube.

4. A bone marrow specimen arrives at the end of a shift and will not be processed for the *Bcl2* translocation until the next day.

Answer: Hold the specimen at refrigeration temperature.

5. The temperature of a refrigerator set at 8°C (±2°C) reads 14°C.

Answer: Recheck the temperature after a few hours while considering alternate refrigeration locations. If the temperature does not return within range, notify the supervisor and relocate sensitive materials to the alternate location.

6. A PCR test for the *BCR/ABL* translocation was negative for the patient sample and for the sensitivity control.

Answer: Repeat the PCR with the addition of a new sensitivity control.

7. A fragile X test result has been properly reviewed and reported.

Answer: Securely file the test results, documents, and associated images electronically or as hard copies in the laboratory archives.

8. A bottle of reagent alcohol with a 3 in the red diamond on its label is to be stored.

Answer: Place the bottle of alcohol in a safety storage cabinet for flammable liquids.

9. The expiration date on a reagent has passed.

   Answer: Discard the reagent. If the material can be used for nonclinical purposes, such as training exercises, label and store the reagent in a separate area away from patient testing reagents.

10. Test results are to be faxed to the ordering physician.

    Answer: Fax the results with a cover sheet containing the proper disclaimers.

Appendix B

Answers to Case Studies

Chapter 10

Case Study 10.1 Interpretation

The woman was successfully engrafted. The recipient peak pattern has converted mostly to the donor peak pattern at 100 days. The percentage of residual recipient cells can be calculated by the analysis of unshared informative alleles at this locus:

$$\%R = \frac{R_{\text{unshared}}}{(R_{\text{unshared}} + D_{\text{unshared}})} \times 100$$
$$= \frac{3{,}171}{(3{,}171 + 40{,}704)} \times 100 = 7.2\%$$

The patient was (100 − 7.2) = 92.8% donor at 100 days, and at 1 year, no residual recipient alleles are present at the level of detection of the instrument (0.5% to 1.0%). At this time, the patient is reported to have more than 99% donor cells and less than 1% recipient cells in the test specimen. The patient is therefore successfully engrafted with donor cells.

Case Study 10.2 Interpretation

These results indicate that the two brothers are identical twins. This means that STR analysis cannot be used for monitoring engraftment after transplant in this case. Advances in epigenetics may lead to discerning molecular characteristics even in identical twins, but not in current practice. With regard to prognosis, there is less chance of graft-versus-host disease in this case, but there is also no graft-versus-tumor effect that can improve the chance of removing all of the recipient tumor cells.

Case Study 10.3 Interpretation

The results show that the material is not of the same genetic origin as that of the patient. Apparently, this microscopic piece of tissue was a contaminant introduced into one of the patient's sections during tissue processing.

Chapter 11

Case Study 11.1 Interpretation

The molecular evidence is consistent with norovirus. Norovirus cannot be cultured. Laboratory tests include electron microscopy, serology, and RT-PCR. Electron microscopy and immune electron microscopy require highly specialized equipment and advanced technical expertise. Interpretation of tests for serum antibodies to the virus is not always straightforward because most of the adult population has serum antibodies to this virus. Furthermore, the development of detectable IgM antibodies may take several days after exposure. RT-PCR, therefore, is the method of choice for detection of this RNA virus. The test targets the viral RNA polymerase gene, which can be used for a broad spectrum of noroviral types. The results in this case showed that the virus was present in the salad lettuce served at the hotel. Lettuce sampled directly from the distributor did not carry the virus, indicating that the contamination occurred at the hotel. This observation was consistent with the discovery of the virus in the hotel workers who prepared the salad. Direct sequencing of cDNA prepared from viral RNA revealed identical sequences for all positive specimens, showing that the guests, workers, and food source shared the same viral strain.

Case Study 11.2 Interpretation

Results from culturing the lysates were consistent with multidrug resistant *Staph aureus* (MRSA). The pulsed field gel electrophoresis (PFGE) analysis revealed that all except one of the isolates from the students were from the same strain. One isolate exhibited two PFGE differences from the others, indicating that it was closely related to these *S. aureus* isolates. Resistance to oxacillin/methicillin results from the expression of an altered penicillin-binding protein encoded by the *mecA* gene. All five isolates shared the type IV *mecA* gene associated with MRSA. The toxin encoded by Panton-Valentine Leukocidin (gene) (PVL), also produced by these isolates, is thought to be responsible for tissue necrosis in MRSA infections.

Further investigation into the cases revealed that all of the affected students had participated in a wrestling meet at one of the high schools. Passage of the organism during this event was the likely source of the shared infections. The meet location was thoroughly cleaned according to the Centers for Disease Control and Prevention recommendations, and the students were encouraged to practice diligent hand washing and optimal hygiene.

Case Study 11.3 Interpretation

The observation that the patient's viral loads were gradually rising was a sign that the virus was developing resistance to the antiviral therapy. Variations of viral quantity within 0.3 log units are considered normal. This patient, however, was seeing significant increases in viral replication over the last 6 months.

Genotyping showed that this patient's virus has a mutation in the reverse transcriptase gene that has rendered the virus resistant to AZT. The patient's treatment should be changed, replacing AZT with another reverse transcriptase inhibitor that would be unaffected by the mutation, such as didanosine or lamivudine. Viral load measurements will be taken regularly to ensure that the change in therapy causes a decrease in viral load over the next few months. Further genotyping may be performed if the viral load starts to trend up again.

Chapter 12

Case Study 12.1 Interpretation

The results in lane 3 show an expansion of the *FMR1* gene promoter, suggesting a premutation in the mother. The full fragile X mutation can be detected by polymerase chain reaction (PCR) and capillary electrophoresis. The cytogenetic results indicate the presence of the fragile X expansion. The size shift caused by the premutation is not detectable by Southern blot. Capillary electrophoresis with triplet-primed PCR produces an extended peak pattern for the premutation. See Figure 12.24B.

Case Study 12.2 Interpretation

The 16,000-bp product is the linearized normal mitochondrial DNA (16,500 bp). The 11,500-bp product is the linearized mitochondrial circle with a 5,000-bp deletion. The deletion is associated with Kearns–Sayre syndrome (KSS), a rare neurological condition affecting muscle function, stature, and hearing, among other symptoms. The results shown are a case of heteroplasmy because both normal and deleted mitochondria are present. The number of deleted mitochondria compared with the number of normal mitochondria will affect the severity of the disease symptoms.

Case Study 12.3 Interpretation

The PCR-restriction fragment length polymorphism (RFLP) results in lane 5 indicate that the patient has the *HFE* C282Y mutation, the most common inherited mutation hemochromatosis. The G → A base change produces an additional recognition site for the *RsaI* restriction enzyme, 30 bp from the end of the 140-bp amplicon in the normal sequence. The absence of the 140-bp fragment in lane 5 shows that the mutation is homozygous (although hemizygosity will also yield this pattern). The symptoms are likely due to iron overload caused by the loss of regulation of iron load through the HFE protein.

Case Study 12.4 Interpretation

The results of the physical and chemistry tests indicate that this patient has a thrombophilic clotting disorder (hypercoagulation). The molecular tests show a homozygous (or hemizygous) factor V Leiden *F5* 1691 G → A mutation. This nucleotide substitution in conjunction with the special primers used for this test will produce a *Hind*III site in the F5 exon 10 PCR product. Results for the prothrombin 20210 G → A mutation are negative (homozygous normal). The mutation in the Factor 5 protein interferes with its ability to block the clotting process, resulting in the symptom of thrombophilia.

Chapter 13

Case Study 13.1 Interpretation

The molecular results revealed a small population of cells in the bone marrow specimen with the t(14;18) translocation. It was apparently below the level of detection of the flow-cytometry and cytogenetics tests for this specimen with minimal B cells. The PCR translocation analysis, however, is lacking amplification controls. There may be a small monoclonal population representing part of the lymphoid aggregates observed in the morphological studies. The complete blood count (CBC) is not representative of this cell population. The immunoglobulin gene-rearrangement results indicate that the gene rearrangement in this small abnormal cell population is not amplifiable or is below the level of detection of this assay. Although detection of the t(14;18) translocation is consistent with follicular lymphoma, the clinical significance of this cell population is not apparent from a single encounter. Patient history and further analyses will provide information as to whether these results are arising from molecular remission or early signs of relapse.

Case Study 13.2 Interpretation

The band pattern in the normal control lane represents the products of the outer primers and one inner primer complementary to the normal *JAK2* sequence. The pattern in lane 3 shows an additional product primed by the *JAK2* V617F mutation–specific primer. The results in lane 2 indicate that the patient does not have the V617F mutation. This test only detects the targeted V617F mutation. It does not rule out the presence of other mutations in *JAK2* exons 12 to 14, which could be assessed as a reflex test, if indicated.

Case Study 13.3 Interpretation

The single-strand conformation polymorphism (SSCP) band pattern in lane 2 is different from that of the normal control in lane 3, indicating the presence of a sequence alteration in the patient DNA. The additional band represents a conformer formed by the folding of the altered sequence. SSCP shows the presence and general location of a sequence variant. Direct sequencing identifies the particular nucleotide change. Synonymous DNA changes do not change the amino acid sequence, and some sequence changes are benign to protein function, so the sequencing information is required for definitive interpretation of the pathogenicity of the sequence change.

Case Study 13.4 Interpretation

The microsatellite instability (MSI) analysis shows instability in at least two of the five microsatellites tested (BAT25 and BAT26). The peak pattern (additional alleles) in the tumor cells compared with normal cells from the same patient results from loss of nucleotides in the repeated sequences. Normally, the losses would be repaired by the mismatch repair system, a protein complex that includes MLH1, MSH2, MSH6, PMS2, and other proteins. If the repair system malfunctions from the loss of one or more of its component parts, the additional alleles will be replicated in ensuing rounds of replication. According to the National Cancer Institute, high MSI is defined by instability in at least two of five NCI-designated microsatellite loci. If more than five microsatellite loci are used to test for MSI, a cut point of 30% of the markers tested is considered MSI. New methods to detect MSI in DNA utilize qPCR or NGS detection of insertions or deletions in other microsatellite sequences.

Case Study 13.5 Interpretation

The t(9;22) translocation is present in 95% of cases of chronic myelogenous leukemia (CML), 25% to 30% of adult acute lymphoblastic leukemia (ALL), and 2% of pediatric ALL. The presence of the translocation is a negative prognostic factor in adult ALL. The t(1;19) translocation may be present in pre-B-ALL. Monoclonal immunoglobulin heavy-chain and light-chain gene rearrangements may be used to monitor B-ALL. Monoclonal T-cell receptor gene rearrangements monitor for T-ALL.

Chapter 14

Case Study 14.1 Interpretation

This patient has DQ alleles, DQA1*05:01 and DQB1*02:01, that form the two parts (alpha chain and beta chain, respectively) of the DQ2 variant heterodimer often found in people with celiac disease. Although a diagnosis cannot be made based on the HLA haplotype alone, the presence of the variant allele is consistent with such a diagnosis made based on clinical and morphological evidence.

Case Study 14.2 Interpretation

A successful transplant does not absolutely require a fully matched donor. In some cases, there may be more risk in delaying a transplant in search of a fully matched donor than in proceeding with the transplant earlier in the disease process. Mismatches at the HLA-B and HLB-C loci may be better tolerated than those at the HLA-A and HLA-DRB1 loci. Donor 3 is the closest match at the HLA-A, HLA-B, and HLA-DRB1 loci. Because donor 3 is homozygous for the A locus, only one determinate is displayed to the recipient immune system, which could lessen the likelihood of graft rejection. Conversely, when the graft does not share one of the recipient alleles, donor T cells may react against the recipient, and graft-versus-host disease may occur. In the present case, however, there is only one mismatch among the tested alleles, and a single mismatch is less associated with graft-versus-host disease (GVHD) than multiple mismatches in the class I and II antigen loci. The actual effect of an allele on the immune response will also depend on the location of the polymorphic amino acid in the protein. A mismatch may be tolerated if the polymorphic amino acid is not exposed on the outer surface of the folded protein.

Case Study 14.3 Interpretation

A parent's alleles will not be the same as the offspring's. Even before typing, the mother will be expected to share at least half of the daughter's alleles. A fully matched sibling is the best donor type. Due to the undefined factors that can affect organ rejection, fully matched siblings are also better than unrelated fully matched donors for successful transplant. In the absence of a fully matched donor, the mother is an acceptable donor for her daughter. Crossmatching results from the daughter's serum were negative for antibodies against antigens displayed on the mother's donated organ, favoring successful transplantation.

Glossary

A

A site adjacent site; location of incoming charged tRNA binding in the ribosome complex for acceptance of the growing peptide

abasic site a position on the DNA sugar-phosphate backbone where the ribose sugar does not carry a nitrogen base

absorptivity constant the characteristic tendency of a substance to absorb light at a given wavelength; used in quantifying nucleic acids, the absorptivity constants of DNA and RNA are (50 ug/mL)/absorbance unit and (40 ug/mL)/absorbance unit, respectively, at 260 nm wavelength

acrocentric having the centromere nearer to one end of the chromosome than the other

adenine one of the common nitrogen bases in DNA and RNA

adenosine one of the common nucleosides in DNA and RNA

affinity maturation selection of B cells expressing antibodies with greater affinity of the antigen

agarose polymer of agarobiose (1,4-linked 3,6-anhydro-α-L-galactopyranose) that, when hydrated, forms a gel frequently used for sieving nucleic acids

alignment comparison and lining up of two or more nucleic acid or protein sequences to achieve the maximal number of identical positions

alkaline phosphatase an enzyme frequently used to generate signals from chemiluminescent or chromogenic substrates

allele a different version of the same sequence, gene, or locus

allele dropout missing sequence data due to loss of DNA fragments during library preparation

allele-specific oligomer hybridization mutation/polymorphism detection technique using immobilized target and sequence-specific probes

allelic discrimination mutation/polymorphism detection using real-time PCR and fluorogenic probes

allelic exclusion expression of a gene on only one of two homologous chromosomes, with the other not expressed

allelic ladder a set of PCR amplicons of a VNTR or STR that represents all possible alleles in a population

alloantibodies antibodies that recognize human antigens

allogeneic genetically from different individuals

allograft a transplanted organ from a genetically different donor

alpha phosphate the phosphate group closest to the ribose sugar in a nucleotide

alpha satellite highly repetitive sequence found at the centromere

alternative splicing removal of introns from RNA using different breakpoints

Alu elements short interspersed nucleotide sequences containing recognition sites for the *Alu* restriction enzyme

ambiguity recognition of two or more antigens by the same antibody

amelogenin locus a gene located on the X and Y chromosome with an XY-specific polymorphism

aminoacyl tRNA synthetases enzymes that covalently attach the appropriate amino acid to its tRNA

amino acid tRNA synthetase one of 20 enzymes that attach the appropriate amino acid to its anticodon-containing tRNA

amino acids nitrogen-containing molecules with specific biochemical properties that are the building blocks of protein

amino-terminal the end of a protein or peptide containing the amino group of the amino acid, usually considered the beginning of the protein's amino acid sequence

ammonium persulfate APS; with TEMED, catalyzes the polymerization of acrylamide gels

amplicon the product of a PCR reaction

amplification replication, copying

amplification control template sequences that are always present; used to confirm the functional competence of the reaction mix

amplification program conditions under which a PCR reaction occurs, including denaturation, annealing, and extension temperatures and times

amplified fragment length polymorphism AFLP; analysis of polymorphic DNA by amplification of selected regions and resolving the amplified products by size

analyte measurement range AMR; the range within which a specimen may be tested directly (without dilution or concentration) for an analyte

analyte-specific reagents ASRs; test components that detect a specific target

analytic accuracy production of correct (true-positive and true-negative) results

analytic sensitivity lower limit of detection of an analyte

analytic specificity ability to detect only the analyte and not nonspecific targets

anaphase stage in mitosis or meiosis where replicated chromosomes separate

ancestral haplotype the haplotype in which a mutation originally occurred

Anderson reference sequence of the mitochondrial hypervariable regions used as a reference for defining polymorphisms

aneuploid having an aberrant number of chromosomes per nucleus, caused by genome mutations

aneuploidy in diploid organisms, having other than two of each chromosome

anion negatively charged atom or molecule

annealing hybridization of complementary sequences

annotation classification of sequence variants based on their biological and/or clinical significance

anode positively charged electrode to which anions migrate

anticipation in genetics, rising phenotypic severity through generations of a family

antigen retrieval protease treatment of tissue to uncover target epitopes

anti-human antibodies AHA; alloantibodies; antibodies that recognize human antigens

antimicrobial agent antibiotic; a substance that inhibits growth of or kills bacteria

antimir a synthetic short RNA that antagonizes oncomirs through complementary hybridization

antiparallel two complementary single-stranded nucleic acids oriented so that when hydrogen-bonded together through complementary bases, the 5′ end of one molecule is next to the 3′ end of the other

antisense strand the single strand of a DNA double helix that is used as a template for messenger RNA synthesis

apoptosis self-directed cell death

arbitrarily primed PCR a PCR method for amplification of complex targets using random priming with 6- to 10-base primers for typing of organisms; also known as randomly amplified polymorphic DNA (RAPD)

arm in cytogenetics, one end of a chromosome, from the centromere to the telomere

array in molecular biology, a set of probes or targets immobilized on the same substrate

autologous genetically from the same individual

autophagy degradation of intracellular substrates within the cell

autoradiograph an image produced by exposing a light-sensitive film to a light- or radiation-emitting membrane

autosomal dominant an inheritance pattern where the child of an affected and an unaffected individual has a 50% probability of being affected

autosomal recessive an inheritance pattern where the child of an affected and an unaffected carrier has a 50% probability of being affected and where the child of two affected individuals has a 100% probability of being affected

autosomal STR short tandem repeats located on other than the X and Y chromosomes

autosome any chromosome except for the X and Y sex chromosomes

avuncular testing determination of the probability of an aunt/uncle–niece/nephew relationship through genetic polymorphisms

B

bacteriocidal kills bacteria

bacteriophage viruses that infect bacteria

bacteriostatic prevents growth of bacteria

balanced in cytogenetics, genetic events that result in no gain or loss of genetic material

balanced polymorphism a DNA sequence difference, the phenotypic effect of which is counteracted by a second trait or polymorphism

band in molecular biology, a pattern on a gel or autoradiogram representing a fragment of nucleic acid or protein

band compression bunching of fragments in gel electrophoresis migration

banding in cytogenetics, the formation of microscopically visible bands on chromosomes by staining

bar coding attachment of short (6-8 base) sample-specific DNA sequences to sequencing templates allowing multiple samples to be sequenced together; also called indexing

Barr body a structure visible in the interphase nucleus formed by the inactive X chromosome in female mammals

basal transcription complex a complex of protein factors required by RNA polymerase to initiate RNA synthesis

base nitrogen base; also used as an expression of units of single-stranded nucleic acid

base extension sequence scanning BESS; mutation/polymorphism scanning using dUTP incorporation into DNA followed by enzymatic cleavage at the dU sites, followed by electrophoretic resolution of the resulting fragments

base pair an expression of units of double-stranded nucleic acid

base pairing association of adenine with thymine/uracil or guanine with cytosine through specific hydrogen bonds

basic local alignment search tool BLAST; a sequence-comparison algorithm

bDNA branched DNA; a signal amplification system

bead array probes attached to fluorescence-labeled beads

beta-2 microglobulin a 12-kD peptide associated at the cell membrane with the MHC heavy chain comprising the class I HLA

beta-lactam a cyclic protein found in antibiotics

beta-lactamase an enzyme produced by bacteria that are resistant to β-lactam–containing antibiotics

betaine N,N,N-trimethyl glycine; used in PCR reactions to increase specificity and yield by facilitating strand separation of the target double helix

bidirectional reading or analysis of both forward and reverse strands of a double-stranded DNA

bin in fragment analysis, a defined range of error expected for electrophoretic migration of the same DNA fragment

binning in fragment analysis, the collection of all peaks or band migration distances within a characteristic distribution in order to determine if two peaks or migration distances are representative of the same allele

bioinformatics an adaptation of information technology for the biological sciences; computational biology

biotin a vitamin that is used in the laboratory for nonradioactive probe detection

bisulfite (DNA) sequencing nucleotide sequence analysis of DNA that has been treated with sodium bisulfite; a method to detect methylated cytosines in DNA

blank a reaction mix containing all components except for target; used to detect target contamination in test reagents

BLAST basic local alignment search tool; a software program that compares nucleotide or protein sequences to find regions of similarity between them

blocking covering nonspecific binding sites before probe hybridization

blot transfer; the membrane carrying the transferred nucleic acid or protein

BOX palindromic repetitive elements found in *Streptococcus pneumoniae*

break-apart probe FISH probes designed to detect translocations where there are multiple possible partners

buffer an ionic substance or solution that maintains specific pH

buffy coat white blood cells separated by density centrifugation

C

C banding centromere staining

calibration fitting an instrument test output with the actual amount of a reference analyte

calibration verification testing of materials of known amounts throughout the measurement range to ensure instrument accuracy

Cambridge reference sequence sequence of the mitochondrial hypervariable regions used as a reference for defining polymorphisms

cancer diseases that are caused by uncoordinated, rapid cell growth

cap 7-methyl G 5′ppp 5′ G/A covalently attached to the 5′ end of messenger RNA

capillary electrophoresis separation of particles in solution inside of a glass capillary

capillary gel electrophoresis separation of particles through a sieving polymer or gel inside of a glass capillary

capillary transfer movement of nucleic acid or protein from a gel matrix to a membrane by capillary action

carboxy-terminal the end of a protein or peptide containing the carboxyl group of the amino acid, usually considered the end of the protein's amino acid sequence

carcinoma tumors of epithelial tissue origin

cathode negatively charged electrode to which cations migrate

cation positively charged atom or molecule

cell division cycle the succession of events when a single cell divides, including pre-replicative phase (G1), DNA synthesis (S), post-replication (G2), and separation (M)

centromere specialized, repetitive DNA sequence that attaches to the spindle apparatus through the kinetochore

centromeric probe CEN, CEP; a probe that binds to centromeres

chain a term sometimes used to describe a polymer of nucleotides comprising a single strand of nucleic acid

chaperone a specialized protein that protects growing peptides as they emerge from the ribosome

charging attachment of an amino acid to tRNA

checkpoint regulated places in the cell division cycle; between the G1 and S phases (G1 checkpoint) and between the G2 and M phases (G2 checkpoint)

chemical cleavage breaking DNA chains with specific nonenzymatic substances

chimera an individual harboring cells or tissue from two different zygotes

chromatin relaxed chromosomes found in interphase nuclei

chromosome a DNA double helix that carries genes

chromosome paint cytogenetic probe analysis designed to visualize an entire metaphase chromosome

chromosome spread an array of intact chromosomes displayed upon hypotonic lysis of a nucleus

class-switch recombination movement of the VDJ gene segment of a rearranged immunoglobulin gene to gene segments encoding IgG, IgE, or IgA constant regions

cleavage-based amplification a signal-amplification system based on the proprietary cleavase enzyme

Clinical Laboratory Improvement Amendments CLIA; recommendations passed by Congress in 1988 to establish quality and testing standards

clinical sensitivity ability of a test result to predict a clinical condition

clinical specificity disease-associated results only in patients who actually have the disease conditions

clone two or more cells or individuals having the same genotype

clonotype the nucleotide sequence of a surface-expressed T-cell receptor gene rearrangement

***coa* typing** tandem repeat element in the 3′ coding region of the *Staphylococcus aureus* coagulase gene, *coa*; analyzed by PFGE to identify MRSA

CODIS Combined DNA Indexing System; a set of 13 STR loci and amelogenin used for positive human identification

codominant offspring simultaneous expression of both parental phenotypes

codon a three-nucleotide sequence that will guide the insertion of a specific amino acid into protein

coefficient of variance CV; a normalized measure of dispersion about the mean; an expression of test precision determined through repeated tests of the same sample analyte

comb in electrophoresis, a device placed in melted agarose to mold wells or spaces for introducing samples to the gel

combined paternity index CPI; the product of the paternity indexes for a set of alleles

combined sibling index likelihood ratio generated by a determination of whether two people share one or both parents; also called sibling index, kinship index

compaction wrapping, folding, and coiling DNA to fit inside the nucleus

comparative genome hybridization CGH; a microarray method to detect deletions or amplifications in DNA

comparative genomic array an array hybridized to DNA in order to detect insertions or deletions

complement a group of proteins that, in combination with antibodies, causes the destruction of particulate antigens

complement-dependent cytotoxicity CDC; killing of cells based on HLA type using reference antigens

complementarity-determining region CDR; a domain in the immunoglobulin heavy-chain gene variable-region coding for amino acids that directly contact antigen

complementary a term used to describe the order of two antiparallel single-stranded nucleic acids such that A on one strand is always across from T on the other strand, and C is always across from G

complementary DNA cDNA; double-stranded DNA synthesized using RNA as a template

computational biology application of statistics, mathematics, and computer science to the study of living systems

conformation-sensitive gel electrophoresis resolution of homoduplexes from heteroduplexes under specified gel conditions

conformer a three-dimensional structure formed by folding and intrastrand hybridization of a single strand of nucleic acid

congenital “born with”; having a genetic component

conjugate a structure comprised of more than one type of molecule, such as a glycoprotein or lipoprotein, or an antibody with covalently attached alkaline phosphatase

conjugated proteins proteins with lipid, carbohydrate, metallic or other components

conjugation transfer of genetic information by physical association of cells

consensus primer a primer directed to the immunoglobulin or T-cell receptor genes with sequences representing the most frequently occurring nucleotide at each position

consensus sequence a family of sequences representing different variations in a population but with similar motifs in nucleotide order

conservation amino acid (or nucleotide) substitutions that preserve the biochemical and physical properties of the original residue or nucleotide sequence

conservative (1) in DNA replication, a term used to describe replication of both strands of the parent double helix simultaneously, producing a daughter double helix comprising two replicated strands; (2) in protein sequences, substitution of an amino acid with one of similar biochemical properties

conservative substitution an alteration in the nucleotide sequence that results in the substitution of an amino acid with one of similar properties

constitutive transcription that is constantly active

contact precautions protective equipment and/or clothing used for direct patient contact and potential exposure to airborne or contact-transmissible infectious agents from the patient

contamination presence of unwanted materials; in PCR, presence of extraneous, nontarget template DNA in a reaction solution

contamination control action taken to avoid or detect unwanted substrates or products of a reaction; in PCR, a reaction mix containing all components except target

continuous allele system genetic structures that vary by increments and are defined by these incremental variations

contract precautions protective equipment and/or clothing used with direct patient contact and potential exposure to airborne or contact-transmissible infectious agents from the patient

control analytes of known type and amount that are included with test specimens to monitor method systems

coverage the number of times a variant is sequenced in a sequencing reaction

CRISPR clustered regularly interspaced short palindromic repeat sequences that provide bacterial immunity to bacteriophage or exogenous plasmids

crossmatching analysis of the reaction between recipient serum antibodies and donor lymphocytes

cross-reactive epitope group CREG; human leukocyte antigens recognized by antibodies that also recognize other HLAs

crRNA short, mature RNAs synthesized from CRISPR regions (processed from pre-crRNA) carrying complementary sequences to a target DNA

cryostat an instrument similar to a microtome that allows cutting slices of frozen tissue at freezer temperatures

cryotube plastic storage tubes designed for storage at freezer temperatures

C_0t value expression of the complexity of DNA sequences

$C_0t_{1/2}$ time required for half of a double-stranded sequence to anneal under a given set of conditions

cut-off value a quantitative assay level that distinguishes positive from negative results; also called cut point

cycle in PCR, one series of extension, annealing, and extension

cycle sequencing semiautomated method to determine the order of nucleotides in a nucleic acid using a thermal cycler

cycling probe a signal-amplification system based on RNase-H digestion of probe bound to target sequences

cytidine one of the common nucleosides in DNA and RNA

cytochrome P-450 a group of enzymes in the endoplasmic reticulum that participates in enzymatic hydroxylation reactions and also transfers electrons to oxygen

cytopathic effect CPE; lysis, death, or growth inhibition of cells due to viral infection

cytosine one of the common nitrogen bases in DNA and RNA

D

DAPI 4′,6-diamidino-2-phenyl indole; DNA-specific dye, used to visualize nuclei

de novo sequencing determination of nucleotide order for the first time

deazaGTP a modified nucleotide used to destabilize secondary structure in template DNA

deletion loss of genetic material

denaturation in nucleic acids, loss of hydrogen bonding between complementary strands; in protein, loss of tertiary structure

denatured alcohol ethanol mixed with other components

denaturing agent in electrophoresis, formamide or urea used to remove secondary structure from molecules

denaturing gel a polyacrylamide gel containing a substance that will separate double-stranded DNA into single strands

deoxyribose ribose without a hydroxyl group on the carbon in the 2 position on the sugar ring

depurination removal of purines from the sugar-phosphate backbone of DNA

derivative chromosome an abnormal chromosome comprising rearranged parts from two or more unidentified chromosomes joined to a normal chromosome

detection limit lower limit of detection of the analyte

dideoxy chain termination a method using dideoxynucleotides to determine the order or sequence of nucleotides in a nucleic acid

dideoxy DNA fingerprinting ddF; mutation/polymorphism screening using one or two dideoxynucleotides and gel electrophoresis

dideoxynucleotide ddNTP; a nucleoside triphosphate lacking a hydroxyl group at the 3′ ribose position

diethyl pyrocarbonate ribonuclease inactivator; DEPC cross-links RNase proteins

digoxygenin steroidal compound used for nonradioactive probe detection

diploid having two of each chromosome

discrete allele system a set of alleles that can be classified by a finite number of types

discriminatory capacity DC; the power of a locus for use in identification, considering the number of different alleles of the locus and the total number of individuals tested in a given population

discriminatory power ability of typing methods to clearly distinguish strains

dispersive a term used to describe DNA replication where both strands of the parent double helix undergo simultaneous, discontinuous replication

dNTP deoxyribose nucleotide triphosphate

domain a functional part of a protein

dominant Mendelian inheritance pattern where offspring of an affected individual and an unaffected mate have a 50% to 100% risk of expressing the affected phenotype

dominant-negative loss of function of multimeric proteins by interaction with one abnormal monomer

dot blot hybridization technique with multiple targets spotted on a membrane all exposed to the same probe

dual-fusion probe dual-color probe; FISH probe designed to detect translocations and their reciprocal products

dye blob artifactual peak pattern caused by residual unincorporated labeled dideoxynucleotides in a sequencing reaction

dye primer covalent attachment of fluorescent molecules to a sequencing primer

dye terminator covalent attachment of fluorescent molecules to dideoxynucleotides

E

E-value in searching for specific sequences, expect value (E) describes the number of sequence matches one can “expect” to see by chance when searching a database of a particular size

eastern blot a hybridization technique used to detect post-translational modification of immobilized proteins

electrode a metallic conductor of an electrical circuit that includes nonmetallic parts

electroendosmosis solvent flow toward an electrode in opposition to particle migration

electrokinetic injection loading of nucleic acid for capillary electrophoresis using positive charges to concentrate the DNA at the end of the capillary

electropherogram data output from capillary electrophoresis where fluorescent signals are recorded as graphical peaks

electrophoresis separation of particles through a solution or matrix under the force of an electric current

electrophoretic transfer movement of nucleic acid or protein from a gel matrix to a membrane using an electric current

end labeling incorporation of radioactive or otherwise labeled nucleotides on the 3′ or 5′ end of a nucleic acid

endonuclease an enzyme that separates DNA or RNA by cutting within the linear molecule

endpoint analysis analysis or observation of a PCR product after the amplification program is complete

engraftment establishment or repopulation of bone marrow cell lineages from a transplanted cell preparation

enzyme induction stimulation of synthesis of RNA-encoding enzymes or other factors from inducible genes

enzyme repression active prevention of the synthesis of RNA-encoding enzymes or other factors from inducible genes

epidemic a disease or condition that affects many unrelated individuals at the same time

epigenetics mitotically and meiotically heritable changes in phenotype not encoded in genotype

episome genetic material within the cell not attached to the host chromosome(s)

epitope specific antigenic sites on a protein

Eppendorf tube plastic conical tubes with 0.5- to 2.0-mL capacity

ERIC enterobacterial repetitive intergenic consensus; palindromic sequences found in gram-negative bacteria; typing of bacteria by analysis of PCR amplicons produced from primers homologous to these sequences

E site position in the ribosome from which the “empty” tRNA is released

ethidium bromide intercalating agent that fluoresces under UV light when bound to double-stranded DNA

euploid having the proper number of chromosomes per nucleus

exclusion in human identification, difference of at least one allele from a reference source

exon sequences of DNA that code for protein

exonuclease an enzyme that removes nucleotides from DNA or RNA from either end of the linear molecule

expression array an array hybridized to RNA in order to measure gene expression

extended haplotype nine STR loci on the Y chromosome, including the eight minimal haplotype loci plus the highly polymorphic YCII locus; used for lineage testing and matching

extended MHC locus xMHC; an 8-Mb region of chromosome 6p including areas flanking the main MHC locus

extracellular outside of the cell

F

F factor a plasmid carrying instructions for the physical association of cells and transfer of genetic information from cell to cell

factor V Leiden a specific mutation in the gene coding for the factor V coagulation protein (F5 1691 A → G) resulting in the substitution of arginine at amino acid position 506 with glutamine

false negative failure to detect an analyte present in a test sample; analogous to Type II or beta error

false positive results suggesting the presence of an analyte that is not in a test sample; analogous to Type I or alpha error

fidelity replication of template sequences without errors

field inversion gel electrophoresis (FIGE) pulsed field gel separation that uses alternation of positive and negative electrodes to resolve large particles

filtering selection of variants found in a sequence based on variant features such as variant frequency, coverage, exon or intron location, germline/somatic status, or other properties

fragment analysis tests based on fragment sizes determined by electrophoretic migration

frameshift insertion or deletion in the nucleotide sequence that throws the triplet code out of frame

framework region FR; a domain in the immunoglobulin heavy-chain gene variable region coding for amino acids that do not directly contact antigen

FRET fluorescent resonance energy transfer; a paired probe detection system for real-time PCR

full chimerism complete replacement of recipient bone marrow with donor bone marrow cells after a transplant

fungicidal having the ability to destroy fungi

fungistatic having the ability to inhibit the growth and reproduction of fungi without killing them

fusion gene a chimeric gene containing parts of two or more separate genes

G

G bands banding patterns of chromosomes after staining with Giemsa

gain-of-function phenotype having a new undesired trait; the new properties of the mutant allele are responsible for the phenotype even in the presence of the normal allele

gamete a haploid reproductive cell

gap discontinuity in one strand of a double-stranded nucleic acid; introduction of spacing used in sequence alignment

GC clamp DNA sequences rich in guanosine and cytosine nucleotides that are more difficult to denature into single strands

gel matrix formed from large hydrated polymers with variably sized spaces through which molecules can move

gel mobility shift assay a method used to detect specific peptides by changes in electrophoretic migration speed upon binding to specific antibodies

gene an order or sequence of nucleotides on a chromosome that contains all the genetic information to make a functional protein or RNA product

gene expression production of RNA and protein using a DNA template

gene rearrangement intrachromosomal deletion and ligation of gene segments in immunoglobulin and T-cell receptor genes

genetic code the relationship between DNA and amino acid sequence

genetic concordance all alleles from two different sources are the same

genome all of the genes in an organism

genomic imprinting enzymatic addition of methyl groups to specific nitrogen bases in a predicted pattern throughout the genome

genotype the genetic DNA composition of an organism

germline having an innate, unmutated, or unrearranged genotype or genetic structure

gonadal mosaicism the presence of more than one genotype in the germ cells of an individual

gradient gel gels with a range of concentrations of urea and polyacrylamide

graft failure inability to detect donor bone marrow cells in a recipient after a bone marrow transplant

graft-versus-host disease GVHD; phenotypic manifestations of immune reaction by donor cells (graft) to the recipient (host) after a bone marrow transplant

graft-versus-tumor effect GVT; immune reaction by donor cells (graft) on residual tumor cells after a bone marrow transplant for treatment of malignant disease

guanine one of the common nitrogen bases in DNA and RNA

guanosine one of the common nucleosides in DNA and RNA

H

hairpin a fold in DNA or RNA that forms a short double strand along the single strand

haploid having one of each chromosome

haplotype a genetically linked set of alleles that are inherited together

haplotype diversity HD; calculated from the frequency of a given haplotype in a defined population

hapten a small molecule capable of causing an immune response when attached to a larger (carrier) molecule

Hardy–Weinberg equilibrium the relative frequency of two alleles in a population is constant

helicase an enzyme that untangles DNA by cutting and ligating one or both strands of the double helix

hematopoietic stem cells HPC; undifferentiated white blood cells that can differentiate into multiple blood cell lineages

hemizygous presence of only one of two possible alleles in a diploid genotype

heteroduplex double-stranded DNA in which the component strands are not completely complementary

heteroduplex analysis HDA; detection of sequence differences by denaturation and renaturation of test and reference double-stranded nucleic acids, forming heteroduplexes where test sequences differ from reference sequences

heterologous noncomplementary; also describes two double-stranded nucleic acids with different nucleotide sequences

heterologous extrinsic control nontarget templates added to a sample before amplification to ensure proper sample purification and amplification

heterologous intrinsic control nontarget templates naturally occurring in a sample used to ensure proper sample purification and amplification

heterophagy intracellular degradation of extracellular substrates taken into the cell by phagocytosis or endocytosis

heteroplasmy mitochondria with different sequences in the same cell

heterozygous in diploid organisms, having different alleles on homologous chromosomes

hierarchical sequencing determination of nucleotide order directed to known regions of the genome

high-density oligonucleotide array a large number of probes (more than 100,000) synthesized in place on the substrate

high-resolution banding staining of less compacted chromosomes to produce a more detailed banding pattern

histocompatible having the same alleles of the MHC locus

histone a basic protein that associates with DNA

histone code the relationship between histone modification (methylation, acetylation, phosphorylation) and control of gene expression

HLA type collection of alleles in the major histocompatibility locus coding for different human leukocyte antigens and detected by genotypic or phenotypic methods

Hoechst 33258 2′-[4-Hydroxyphenyl]-5-[4-methyl-1-piperazinyl]-2,5′-bi-1H-benzimidazole trihydrochloride pentahydrate; DNA-specific dye used in fluorometric quantification of DNA

holoenzyme a multisubunit protein with enzymatic function

home-brew describing a test system developed and validated within the clinical laboratory

homoduplex double-stranded DNA in which the component strands are completely complementary

homologous complementary; also describes two double-stranded nucleic acids with the same nucleotide sequence

homologous extrinsic control a PCR template with primer-binding sites matching test targets and a nontarget insert

homoplasmy all mitochondria in a cell having the same sequences

homozygous in diploid organisms, having the same allele on both homologous chromosomes

hot-start PCR preparation of PCR reaction mixes so that no enzyme activity is possible until after the mixes have gone through an initial heating in the thermal cycler

HPLC high-performance liquid chromatography or high-pressure liquid chromatography; separation of components of a mixture in solution using a column packed with stationary-phase material packed at high pressure

human leukocyte antigens HLA; gene products of the major histocompatibility locus in humans

humoral sensitization development of anti-human antibodies, often due to organ transplant, blood transfusion, or pregnancy

hybrid a product of genetically unrelated parents

hybrid capture an ELISA-like assay using antibodies to an RNA:DNA hybrid formed by hybridization of target RNA with a DNA probe

hybrid resistance rejection of transplanted organs by immunocompromised mice

hybrid vigor survival advantage observed in offspring of genetically unrelated parents over offspring of genetically similar parents

hybridization in nucleic acids, the formation of hydrogen bonds between complementary strands of DNA or RNA

hybridoma hybrid cells, often used to make monoclonal antibodies

hydrogen bond shared electrons between hydrogen atoms; the basis of base pairing in nucleic acids

hypervariable region HV; a region found to have many different alleles (sequences) in a population; HV I and HV II are two such regions in mitochondrial DNA

I

iatrogenic infection caused by the actions of a physician

immobilized in molecular biology, bound to a substrate such as a nitrocellulose-based membrane or glass slide

imprinting marking of DNA by the enzymatic addition of methyl groups to specific nitrogen bases

in silico analysis performed on a computer or by computer simulation

in vitro analytical tests IVATs; reagent sets approved by the FDA only for the detection of specific analytes with no claim to clinical utility

in vitro diagnostic IVD; reagent sets intended for use in the diagnosis of specific diseases; these may be FDA-approved tests

inclusion in human identification, concordance of alleles with a reference source

indel an insertion in DNA with a concomitant deletion

indexing attachment of short (6-8 base) sample-specific DNA sequences to sequencing templates allowing multiple samples to be sequenced together; also called bar coding

inducible transcription that is regulated by proteins and other factors

informative locus a genetic locus that differs between two individuals

insertion gain or duplication of genetic material

internal control an analytical target distinguishable from the test target but detectable with test reagents that is included in the reaction mix with the test target

internal labeling incorporation of radioactive nucleotides in chain termination sequencing for visualization of the sequencing ladder

internal transcribed spacer ITS; conserved elements found in regions separating the ribosomal RNA genes; used for typing yeast and molds

interphase cell state in between mitosis or meiosis

interphase FISH cytogenetic probe analysis of mitotic nuclei

intracellular inside of the cell

intron sequences in DNA that interrupt coding sequences

Invader assay mutation/polymorphism detection using sequence-specific probes and a proprietary cleavase enzyme

inversion intrachromosomal breakage, reversal, and reunion of genetic material

inversion probe mutation/polymorphism detection method based on sequence-specific probe hybridization, extension, ligation, and amplification

investigational use only IUO; reagents not intended for use on patient samples

isochromosome chromosome containing two copies of the same arm and loss of the other arm

isocratic a mobile phase in chromatography that remains at constant concentration throughout the column

isodecoders transfer RNAs sharing anticodon sequences but differing in the sequences outside of the anticodon

isotype a protein related to another through common function or structure; immunoglobulin classes directed against the same antigen

K

kappa-deleting element KDE; a DNA sequence that determines deletion of the IgK constant region in cells producing the IgL light chains

karyotype the complete set of metaphase chromosomes of a cell

killer cell immunoglobulin-like receptor KIR; a protein expressed on the surface of NK cells and some memory T cells that interacts with HLA receptors

kinetochore a protein structure that attaches the centromeres of chromosomes to the spindle apparatus during cell division

kinship index likelihood ratio generated by a determination of whether two people share one or both parents; also called sibling index, combined sibling index

Klenow fragment the large fragment of DNA polymerase without its exonuclease activity

L

laboratory-developed tests LDT; reagent sets and procedures validated in individual laboratories and not subject to special controls by the FDA

lagging strand one strand of the parent double helix that is read 5′ to 3′ and replicated discontinuously

leading strand one strand of the parent double helix that is read 3′ to 5′ and replicated continuously

leucine zipper specialized secondary structure found in proteins that bind DNA with leucine residues at each seventh position for 30 residues

leukemia neoplastic disease of blood in which large numbers of white blood cells populate the bone marrow and peripheral blood

leukocyte receptor cluster a group of genes on the long arm of chromosome 19 coding for the killer cell immunoglobulin-like receptors

library a set of short DNA fragments, prepared from genomic DNA by nuclease digestion and amplification or amplification with multiple PCR primers

ligase an enzyme that forms one phosphodiester bond connecting two preexisting fragments of DNA

likelihood ratio comparison of the probability that two genotypes come from the same source with the probability that they come from different sources

limit of detection detection limit; lower limit of detection; the lowest target concentration that can be detected 95% of the time in a test assay

linearity quantitative correlation between test result and actual amount of analyte

linkage analysis use of alleles or phenotypes from known locations in the genome to map genes

linkage disequilibrium the assumption that two alleles are genetically linked

linkage equilibrium the assumption that two alleles are not genetically linked

linker region DNA between histones

locked nucleic acid LNA; nucleic acid with modified sugar-phosphate backbone

locus a defined site or location in a genome

locus genotype the set of inherited alleles at a particular site on homologous chromosomes, such as STR repeats or SNPs

locus-specific brackets definition of a high and low limit of migration within which alleles are identified with a given degree of certainty

locus-specific RFLP detection of polymorphisms in restriction sites within designated regions or genes; used for typing of microorganisms

logit analysis a statistical method intended to translate a quantifiable measure to a predicted state or event (e.g., level of gene expression and clinical state)

long interspersed nucleotide sequences LINEs; highly repeated DNA sequences of 6 to 8 kbp in length located throughout the human genome

loss-of-function phenotype lacking a desired trait; in diploid organisms, inactivation of the normal allele is responsible for the phenotype

loss of heterozygosity LOH; deletion or inactivation of a functional allele, leaving a mutated allele

lymphoma neoplasm of lymphocytes, capable of forming discrete tissue masses

Lyon hypothesis only one X chromosome remains genetically active in females; also called Lyon's hypothesis

lysosome cellular organelle in which cellular products are degraded

M

M13 a single-stranded DNA bacteriophage used in early procedures to make single-stranded templates for sequence analysis

M13 universal primer a sequencing primer that could be used for all sequences cloned into the M13 RF plasmid

macroarray a membrane containing multiple immobilized probes

major groove the larger cleft in the double-stranded DNA helix

major histocompatibility locus MHC; a group of genes located on chromosome 6p in humans

marker a site (gene or polymorphism) of known location in a genome

marker chromosome an unknown chromosome or part of a chromosome of unknown origin

master mix preassembled components of a reaction solution

match identity or inheritance of one or a set of genetic alleles

match probability probability of identity or inheritance of a set of genetic alleles

matched unrelated donor MUD; a donor selected from a pool based on HLA typing for a transplant

Maxam–Gilbert sequencing a chemical sequencing method based on controlled breakage of DNA

mecA The gene encoding the altered penicillin-binding protein, PBP2a or PBP2′, which has a low binding affinity for methicillin

meiosis replication and separation of DNA by reductional and equational divisions in diploid organisms, resulting in four haploid products

melt-curve analysis mutation/polymorphism scanning based on changes in hybridization temperature

melting temperature T_m; the temperature at which DNA is 50% single stranded and 50% double stranded

messenger RNA ribonucleic acid that carries information from DNA in the cell nucleus to ribosomes in the cytoplasm

metacentric having the centromere in the middle of the chromosome

metaphase mitosis or meiosis phase where replicated compacted chromosomes align and prepare to separate

metaphase FISH cytogenetic probe analysis of metaphase chromosomes

metastasis movement of tumor cells from the original (primary) site of the tumor to other locations

microarray a small glass slide or other substrate containing multiple immobilized probes

microdeletion loss of a small amount of genetic material barely detectable or undetectable by karyotyping

microelectronic array a small substrate containing multiple immobilized probes subject to electrodes at each position on the array

microfuge a small centrifuge designed to accommodate 0.2- to 2.0-mL tubes

microsatellite DNA from highly repeated short sequences (STRs) such that it differs in density from nonrepeated DNA

microsatellite instability MSI; contraction and expansion of mononucleotide and dinucleotide repeat sequences in DNA caused by lack or repair of replication errors

microvariant a repeated sequence in DNA missing part of the repeat unit

migration in electrophoresis, movement of particles through a matrix under the force of a current

minimal haplotype eight Y-STR loci used for paternal lineage testing and matching

minimum inhibitory concentration MIC; the least amount of antimicrobial agent that inhibits the growth of an organism

minisatellite DNA from highly repeated sequences (VNTRs) such that it differs in density from nonrepeated DNA

mini-STR a system of PCR amplification of an STR locus using primers that produce a smaller amplicon than standard STR systems

minor groove the smaller cleft in the double-stranded DNA helix

minor groove binding MGB; the nature of association of a molecule with double-stranded DNA within the minor groove of the double helix

minor histocompatibility antigens mHags; proteins outside of the main MHC locus that affect organ engraftment

misprime aberrant initiation of DNA synthesis from a primer hybridized to template sequences different from the intended target

mitochondrion a subcellular organelle responsible for energy production in the cell

mitogen an agent that promotes cell division

mitosis separation of DNA in diploid organisms by equational division, resulting in two diploid products

mixed chimerism partial replacement of recipient bone marrow with donor bone marrow cells after a transplant

mixed lymphocyte culture MLC; measurement of growth of lymphocytes activated by serum antibodies as an indication of donor/recipient incompatibility; co-culturing of lymphocytes from unrelated individuals to determine HLA type

mobility capacity for migration

Molecular Beacons a self-annealing probe detection system used in real-time PCR

monoclonal (monotypic) having the same genotypic or phenotypic composition

monoclonal antibody antibody preparation designed to recognize a defined antigen or epitope

monomer a single protein or peptide that can function alone

monosomy in diploid organisms, loss of a single chromosome

mosaic an individual harboring genotypically different cells or tissue arising from a single zygote

MRSA methicillin-resistant *Staphylococcus aureus*; a strain of *S. aureus* that has resistance to multiple antibiotics

multilocus sequence typing MLST; characterization of bacteria using sequences of internal fragments of housekeeping genes

multiple-locus probe MLP; a Southern blot probe set that hybridizes to more than one locus on the same blot

multiplex PCR amplification of more than one target in a single PCR reaction

mutation a change in the order or sequence of nucleotides in DNA found in less than 1% to 2% of a given population

myeloablative in bone marrow transplants, removal of recipient bone marrow

N

NASBA nucleic acid sequence–based amplification; an isothermal DNA amplification process that includes an RNA intermediate

necrosis cell and tissue destruction upon cell death

negative control an analytical target that does not react with reagent/detection systems; used to demonstrate that a procedure is functioning with proper specificity and without contamination

negative template control template sequences that do not match the intended target

neoplasm growth of tissue that exceeds and is not coordinated with normal tissue; a tumor

nested PCR amplification of the same target in two PCR reactions, using the product of the first amplification as the template for a second round of amplification with primers designed to hybridize within the amplicon formed by the first primer set

new mutation a spontaneous mutation arising in germ cells of an unaffected individual

nick translation addition of labeled nucleotides at nicks (single-strand breaks) in DNA

nitrocellulose matrix material with high binding capacity for nucleic acids and proteins

nitrogen base a carbon-nitrogen ring structure that comprises part of a nucleoside

nonconservative substitution an alteration in the nucleotide sequence that results in the substitution of an amino acid with one of dissimilar properties

nondisjunction abnormal separation of chromosomes during cell division resulting in both of a chromosome pair in one daughter cell

nonhistone proteins chromosome-associated proteins that maintain chromosomal compaction in the nucleus

noninformative locus a genetic locus that is identical in two individuals

nonisotope RNase cleavage assay NIRCA; a mutation/polymorphism screening and amplification method using RNase cleavage of heteroduplexes

nonpolar hydrophobic; water insoluble

nonsense codon a trinucleotide sequence that terminates translation; UAA, UAG, UGA

nonsense-mediated decay degradation of mRNA with premature stop codons

nonsense substitution an alteration in the nucleotide sequence that results in the substitution of a termination codon for an amino acid codon

northern blot a method used to analyze specific RNA transcripts in a mixture of other RNA

nosocomial hospital-caused infection

NTP nucleotide triphosphate

nucleic acid a linear polymer of nucleotides

nucleic acid sequence-based amplification NASBA: replication of a DNA or RNA target through an intermediate RNA product

nuclein viscous substance isolated from nuclei by Miescher, later to be known as DNA

nucleoside a unit of nucleic acid comprised of a ribose sugar and a nitrogen base

nucleosome eight histones wrapped in approximately 150 bp of DNA

nucleotide a unit of nucleic acid comprising a phosphorylated ribose sugar and a nitrogen base

null allele a genetic type or variant with no phenotypic effect

O

Okazaki fragments intermediate products of DNA synthesis on the lagging strand

oligoclone a small subpopulation with identical genotype or phenotype within a larger group

oligomer proteins or peptides that function as components of more than one unit (dimer, trimer, tetramer); in nucleic acids, a short sequence of nucleic acid

oligonucleotide array a small glass slide or other substrate containing multiple immobilized probes synthesized directly on the substrate; high-density oligonucleotide arrays

oncogene a mutated gene that promotes the proliferation and survival of cancer cells

oncology the study of cancer

oncomir a microRNA that acts as a tumor-suppressor gene or an oncogene

ontogeny the course of development, for example, of an individual organism

Oxford sequence sequence of the mitochondrial hypervariable regions used as a reference for defining polymorphisms

P

p in cytogenetics, indicating the short arm of a chromosome

P site peptidyl site, location of charged tRNA binding in the ribosome complex

PAM protospacer adjacent motif; a short DNA sequence required for recognition and cutting of target sequences by Cas nucleases

pandemic a disease or condition found across wide geographical areas at the same time

panel reactive antibodies PRA; a set of reference antibodies used to define the HLA of reference lymphocytes

paracentric inversion intrachromosomal breakage, reversal, and reunion of genetic material, not including the centromere

paternity index an expression of how many times more likely it is that a child's allele is inherited from an alleged father than from a random man in the population

pedigree a diagram of the inheritance pattern of a phenotype (or genotype) in a family

penetrance frequency of expression of a disease phenotype in individuals with a gene lesion

peptide a polymer of a few amino acids

peptide bond covalent carbon-nitrogen bonds that connect amino acids in proteins

peptide nucleic acid PNA, nucleic acid with peptide bonds replacing the sugar-phosphate backbone

peptidyl transferase ribozyme that catalyzes the formation of a peptide bond

pericentric inversion intrachromosomal breakage, reversal, and reunion of genetic material, including the centromere

pH drift decrease of pH at the cathode and increase in pH at the anode during electrophoresis

phenotype the biological properties of an organism

phosphodiester bond covalent attachment of the hydroxyl oxygen of one phosphorylated ribose (or deoxyribose) sugar to the phosphate phosphorous of the next

phosphor a substance that glows after exposure to electrons or ultraviolet light

photobleaching fading; photochemical destruction of a fluorescent molecule (fluorophore) by exposure to light

Phrap a sequence assembly tool

Phred a sequence quality assessment system

phylogeny history and evolutionary development of a species or related group of organisms

plasma cells mature white blood cells that produce antibodies

plasmid small, usually circular double-stranded DNA, often carrying genetic information, that replicates independently or in synchrony with host cell replication

plasmid fingerprinting isolation and restriction mapping of bacterial plasmid for epidemiological studies

point mutation DNA sequence alteration involving one or a few base pairs

polar hydrophilic; water soluble

polarity orientation of nucleic acid such that one end (5′) has a phosphate group and the other end (3′) has a hydroxyl group and is the site of extension of the molecule

polony a cluster of PCR products primed by surface-attached oligonucleotides

polyA tail polyadenylic acid at the 3′ terminus of messenger RNA

polyacrylamide polymer of acrylamide and bis-acrylamide that, when hydrated, forms a gel frequently used for sieving proteins and nucleic acids

polyadenylate polymerase the enzyme that catalyzes the formation of polyadenylic acid at the 3′ terminus of messenger RNA

polyadenylation addition of adenosines to the 3′ end of a messenger RNA molecule

polyadenylation signal the site of activation of termination in eukaryotic RNA synthesis

polyadenylic acid nucleic acid comprising only adenosine nucleotides

polyclonal (polytypic) having a variety of genotypes or phenotypes

polyclonal antibody a collection of antibodies produced in vivo that recognizes an antigen or epitope

polymerase in nucleic acids, an enzyme that connects nucleotides by catalyzing the formation of phosphodiester bonds

polymerase chain reaction PCR; primer-directed DNA polymerization in vitro resulting in amplification of specific DNA sequences

polymorphism a difference in DNA sequence found in 1% to 2% or more of a given population

polypeptide protein

polyphred a sequence quality assessment designed to detect single-nucleotide changes

polyploidy in diploid organisms, having more than two chromosome complements

polysomy in diploid organisms, having more than two of a single chromosome

polythymine nucleic acid comprising only thymine nucleotides; poly T

polyuracil nucleic acid comprising only uracil nucleotides; poly U

position effect differences in gene transcription depending on chromosomal location

positive control an analytical target known to react with reagent/detection systems; used to demonstrate that a procedure is functioning properly

post-PCR laboratory areas and materials used for working with PCR products

preanalytical error events occurring prior to sample analysis that adversely affect test results

precision reproducibility of test results

prehybridization treatment of membranes before introduction of probe to minimize nonspecific binding

premutation a genetic lesion not manifesting a disease phenotype but prone to advance to full mutation/disease status

pre-PCR laboratory areas and materials not exposed to PCR products

preventive maintenance routine review of instrument function with appropriate part replacement

primary antibody antibody that directly binds a target molecule

primary resistance mutation mutations that are specific for a given antibiotic or antiviral drug and make an organism less susceptible to that drug

primary structure in proteins, the sequence of the amino acids; in nucleic acids, the sequence of nucleotides

primase the enzyme that produces short (~60 bp) complementary RNAs to prime DNA synthesis

primer a short, single-stranded DNA fragment designed to primer synthesis of a specific DNA region

primer dimer a side product of PCR resulting from amplification of primers caused by 3′ homology in primer sequences

prion aberrantly folded proteins capable of transferring their configuration to normal proteins

prior odds initial probability of an event in the absence of a test effect

private antibodies HLA type-specific antibodies

probability of paternity a number calculated from the combined paternity index and reported in paternity testing

proband the initial patient resulting in a genetic study of a family

probe in blotting procedures, nucleic acid or protein with a detectable signal that specifically binds to complementary sequences or target protein

probe amplification an approach to increase the sensitivity of target detection by protection and amplification of the probe, rather than the target

probit analysis a statistical method to form binary groups based on quantitative data; used to determine cut points for qualitative tests

processivity the ability of polymerase to stay with the template during replication of long sequences

product rule the frequency of a genotype in a population is the product of the frequencies of its component alleles

proficiency testing analysis of specimens supplied to independent laboratories from an external reference source

profile a set of markers, alleles, or other characteristics in an individual

promoter DNA sequences that bind RNA polymerase and associated factors

proofreading a correction function of the DNA polymerase complex that repairs errors committed during DNA replication

prophase early mitosis or meiosis

prosthetic group nonprotein moiety, such as a sugar or lipid group covalently attached to a protein

protease an enzyme that dissembles protein into its component amino acids

protein a polymer of amino acids with structural or functional capabilities

protein truncation test mutation/polymorphism analysis using in vitro transcription–translation systems

proteome all of the proteins that make up a cell

proteomics study of the proteome, usually through array or mass spectrometry analysis

proto-oncogene a normal gene with the potential to promote aberrant survival and proliferation if mutated

protospacer the part of the CRISPR crRNA sequence that is complementary to the target

pseudoautosomal describing homologous regions in the X and Y chromosomes that undergo crossing over during meiosis

pseudogene intron-less, nonfunctional copies of genes in DNA

psoralen(s) a substance from plants that covalently attaches to DNA under ultraviolet light

public antibodies antibodies that recognize more than one HLA type

pulsed-field gel electrophoresis PFGE, a technique using alternating currents to move very large DNA fragments through an agarose gel

purine nitrogen bases with a double-ring structure; adenine and guanine are purines

pyrimidine nitrogen bases with a single-ring structure; cytosine, thymine, and uracil are pyrimidines

pyrimidine dimers boxy structures resulting from covalent attachment of cytosine or thymine with an adjacent thymine in the same DNA strand; formed when DNA is exposed to UV light

pyrogram luminescence data output from a pyrosequencer

pyrophosphate an ester of pyrophosphoric acid containing two phosphorous atoms, $P_2O_7^{4-}$

pyrosequencing a sequencing method based on the release of pyrophosphates during DNA replication

Q

q in cytogenetics, indicating the long arm of a chromosome

Q banding banding patterns of chromosomes after staining with quinacrine fluorescent dyes

Qβ replicase RNA-dependent RNA polymerase from the bacteriophage Qβ

quantitative PCR qPCR, real-time PCR

quaternary structure functional association of separate proteins or peptides

quencher a molecule that takes fluorescent energy from a reporter dye

R

R banding modified G banding using acridine orange; the resulting pattern will be opposite of that of G banding

R loop a hybrid structure formed between RNA and DNA such that the unpaired introns in the DNA form loops

random priming template-directed synthesis of single strands of DNA carrying labeled nucleotides

randomly amplified polymorphic DNA RAPD; a PCR method using random priming with 6- to 10-base primers for typing of organisms; also known as arbitrarily primed PCR

reagent alcohol a mixture of 90.25% ethanol, 4.75% methanol, and 5% isopropanol

reagent blank PCR reaction mix with all components except template

real-time PCR PCR performed in an instrument that can measure the formation of amplicons in real time; also called quantitative PCR (qPCR)

recessive Mendelian inheritance pattern where offspring have a 25% likelihood of expressing an affected parental phenotype

reciprocal translocation exchange of DNA between separate chromosomes with no gain or loss of genetic material

recognition site sequences in DNA that are bound by proteins such as restriction enzymes or transcription factors

recombinant a new combination of DNA sequences that is produced from different preexisting (parent) sequences; a cell or organism containing a new combination of sequences

recombine to mix different DNA sequences such that a new combination of sequences is produced from different preexisting (parent) sequences

redundant tiling design of probes on an array such that a predictable mutation is located at different places in the probe sequence

reference range expected analyte frequency or levels from a population of individuals

REF-SSCP mutation/polymorphism scanning by SSCP in combination with restriction enzyme fingerprinting; see single-strand conformation polymorphism

regulatory sequence that part of a gene sequence that binds factors controlling the expression of the gene

release factor specialized protein required for the dissociation of the translation complex at the end of protein synthesis

reportable range the range within which test results are considered to be valid (with or without dilution)

reporter dye a fluorescent dye whose signal indicates the presence of target

REP-PCR repetitive extragenic palindromic PCR; analysis of amplicons produced from primers homologous to interspersed recurring sequences in microorganisms

reproducible consistency of test results produced from the same procedure

research use only RUO; reagents not intended for use on patient samples

resequencing direct sequence analysis of the same area in multiple samples to detect variants; usually applied to known mutations or areas of frequent mutation

resolution the level of detail at which an allele is determined; also, the separation of particles by electrophoresis or chromatography

restriction endonuclease an enzyme that makes double-stranded cuts in DNA at specific sequences

restriction enzyme an endonuclease isolated from bacteria that will break single- or double-stranded nucleic acids at specific sequences

restriction fragment length polymorphism RFLP; a sequence variation that results in creating, destroying, or moving a restriction site

restriction map a diagram showing the linear placement of restriction enzyme recognition sites in DNA

reverse dot blot hybridization technique with multiple probes spotted on a membrane all exposed to the same target

reverse transcriptase an RNA-dependent DNA polymerase

reverse-transcriptase PCR RT-PCR; polymerase chain reaction amplifying cDNA made from RNA by reverse transcriptase

RF part of the M13 life cycle, where its genome is in the form of a double-stranded DNA circle

ribonuclease an enzyme that separates RNA into its component nucleotides

ribonucleic acid RNA; a polymer of ribonucleotides

ribose a five-carbon sugar

ribosomal binding site mRNA sequence that binds to ribosomes in preparation for protein synthesis

ribosomal RNA ribonucleic acid that serves as a structural and functional component of ribosomes

ribosome complex of RNA and proteins that catalyzes formation of peptide bonds

ribotyping typing of microorganisms using restriction fragment polymorphisms in ribosomal RNA genes

ribozyme RNA with enzymatic activity

ring chromosome a circular structure formed from the fusion of the ends of a chromosome or chromosome fragment

RNA integrity control an amplification control included in RT-PCR to distinguish negative and false-negative results

RNA-dependent RNA polymerases enzymes that synthesizes RNA using RNA as a template

RNA polymerase an enzyme that catalyzes the template-dependent formation of phosphodiester bonds between ribonucleotides forming ribonucleic acid

RNAse-free (RNF) reagents, consumables, and facilities prepared or treated to protect RNA from RNA digesting enzymes

RNase H an exonuclease that digests RNA from RNA:DNA hybrids

RNase protection a method used to detect transcripts by hybridizing test RNA with a labeled probe and removing un-hybridized target and probe with single-strand-specific nucleus

RNA-SSCP rSSCP; mutation/polymorphism scanning using RNA instead of DNA; see single-strand conformation polymorphism

S

S sedimentation coefficient; used to describe the density (movement) of particles in a given medium under gravity or centrifugal force

S1 analysis a method used to detect transcripts by hybridizing test RNA with a labeled probe and removing un-hybridized target and probe with single-strand-specific nucleus

salting out purification of nucleic acid by precipitating proteins and other contaminants with high salt at low pH

Sanger sequencing dideoxy chain termination sequencing

sarcoma tumor of bone, cartilage, muscle, blood vessels, or fat

Scorpion a self-annealing labeled primer detection system for real-time PCR

secondary antibody antibody that binds primary antibody to a target molecule

secondary resistance mutation secondary mutations that enable an organism with primary resistance mutations to replicate in the presence of a given antibiotic or antiviral drug

secondary structure specific folding of protein or nucleic acid through the interaction of amino acid side chains or nitrogen bases, respectively

selenocysteine cysteine with selenium replacing the sulfur atom; the 21st amino acid

selenoprotein a protein containing selenocysteine

self-sustaining sequence replication SSSR or 3SR; replication of a DNA or RNA target through an intermediate RNA product

semi-conservative a term used to describe DNA replication where one strand is conserved and serves as a template for a new strand, resulting in a new double helix comprising one parent and one daughter strand

semi-nested PCR amplification of the same target in two PCR reactions, using the product of the first amplification as the template for a second round of amplification with primers, one of which is the same as the first round and one of which is designed to hybridize within the amplicon formed by the first primer set

sense strand the single strand of a DNA double helix that is not used as a template for messenger RNA synthesis

sensitivity control an analytical target known to react with reagent/detection systems but present in the lowest detectable concentration; used to demonstrate that the procedure is detecting target down to the indicated input levels of testing

sequence the order of nucleotides in a nucleic acid polymer; the order of amino acids in a protein polymer

sequence-based typing SBT; determination of HLA DNA polymorphisms by direct sequencing

sequence complexity the length of nonrepetitive nucleic acid sequences

sequence-specific oligonucleotide probe hybridization SSOP, SSOPH; detection of DNA polymorphisms using amplified test DNA immobilized on a filter exposed to labeled probes that will hybridize to specific sequences

sequence-specific (primer) PCR SSP, SSP-PCR; detection of point mutations or polymorphisms in DNA with primers designed so that the 3′-most base of the primer hybridizes to the test nucleotide position in the template

sequencing ladder the pattern of sequence reaction products on a polyacrylamide gel

serology study of antigens or antibodies in serum

sex-linked having genetic components located on the X or Y chromosome

sgRNA synthetically fused crRNA and tracrRNA that can be designed to target specific DNA sequences for cutting with Cas9 endonuclease

sharkstooth comb a specialized comb used to generate lanes immediately adjacent to one another to facilitate lane-to-lane comparison

short interspersed nucleotide sequences SINEs; highly repeated sequences approximately 0.3 kbp in length located throughout the human genome

short tandem repeats STR; head-to-tail repeats of DNA sequences with less than 10-bp repeat units

shotgun sequencing determination of nucleotide sequence by the assembly of sequence data from a random collection of small fragments

sibling index likelihood ratio generated by a determination of whether two people share one or both parents; also called kinship index or combined sibling index

side chain that portion of an amino acid with biochemical properties

signal amplification an approach to increase the sensitivity of target detection by amplification of the probe signal rather than the target

signal transduction transfer of extracellular and intracellular stimuli through the cell cytoplasm to the nucleus, ultimately affecting gene-expression patterns

silent substitution an alteration in the nucleotide sequence that does not change the amino acid sequence

silver stain a sensitive staining system for nucleic acids and proteins after gel electrophoresis

single-locus probe a Southern blot probe that hybridizes to only one locus

single-nucleotide polymorphism SNP; a 1-bp variation from a reference DNA sequence

single-strand conformation polymorphism SSCP; a method designed to detect sequence alterations in DNA through differences in secondary structure due to differences in the nucleotide sequence

SISH silver-enhanced in situ hybridization; a bright-field hybridization method similar to CISH for the detection of chromosomal abnormalities and gene amplifications

slot blot a membrane containing multiple targets deposited in an elongated pattern rather than a “dot”

solution hybridization hybridization of nucleic acid target and labeled probe in solution

somatic hypermutation enzymatic alterations of nucleotide sequences in the variable region of the immunoglobulin heavy-chain gene

Southern blot a method used to analyze specific regions of DNA in a mixture of other DNA regions

***spa* typing** tandem repeat elements in the 3′ coding region of the *Staphylococcus aureus* protein A gene, *spa,* analyzed by PFGE to identify MRSA

spectral karyotyping cytogenetic probe analysis designed to visualize the entire complement of metaphase chromosomes, with unique colors for each chromosome

spin column small silica columns designed to fit within an Eppendorf tube, using centrifugal force to wash and elute bound substances

splicing Removal of intron sequences from messenger RNA

split chimerism differential replacement of specific recipient bone marrow cell lineages with donor bone marrow cells after a transplant

split specificity antigenic types derived from a parent antigen

standard curve graphical depiction of results from a dilution series of analyte encompassing levels predicted to result from unknown test specimens

standard precautions recommendations by the Centers for Disease Control and Prevention for handling potentially infectious specimens

standard tiling design of probes on an array such that the mutation site is always in the same position from the end of the probe

Stoffel fragment *Taq* polymerase lacking 289 N-terminal amino acids

stop codon nonsense codon

strand one molecule of nucleic acid; a double helix contains two strands of DNA

strand displacement amplification SDA; an isothermal amplification reaction using nick translation and strand displacement

structural sequence that part of a gene sequence that codes for amino acids

stutter a PCR artifact that results from slippage of the template on short repeated sequences

submarine gel agarose gels loaded under the buffer surface

submetacentric describing a chromosome with a centrally located centromere closer to one end of the chromosome than the other

subpopulation genetically defined groups within a population

sugar-phosphate backbone the chain of phosphodiester bonds that hold nucleotides together in a single strand of nucleic acid

surname test a test of paternal lineage using alleles on the Y chromosome

susceptibility testing detection of resistance to given antimicrobial agents by the ability to grow in the presence of the agent

SYBR green a fluorescent dye specific to double-stranded nucleic acid

synonymous describing DNA sequence variants that do not change the reference amino acid sequence

T

tag SNP one or two single-nucleotide polymorphisms that can be used to define a haplotype or ~10,000-bp block of DNA sequences

***Taq* polymerase** a heat-stable polymerase isolated from *Thermus aquaticus*

TaqMan a probe detection system used in real-time PCR

target specific sequences or regions that are detected in an analytical procedure

target amplification replication of a specific target region of DNA

TE buffer 10-mM Tris, 1-mM EDTA buffer

telocentric having the centromere at one end of the chromosome

telomere the end of a chromosome; composed of repeated DNA sequences and associated proteins

telomeric probe probes that bind to telomeres

TEMED N,N,N′,N′-tetramethylethylenediamine, with APS, catalyzes polymerization of acrylamide gels

template a region of single-stranded nucleic acid used as a guide for replication or synthesis of a complementary strand

teratocarcinoma tumors comprising multiple cell types

terminal transferase an enzyme that adds nucleotides to the end of a strand of nucleic acid without using a template

tertiary structure configuration of folded proteins required for function

tetraploid having four of each chromosome

threshold cycle C_T; in real-time PCR, the PCR cycle at which the fluorescence from the PCR product surpasses a set cut point

thymidine one of the common nucleosides in DNA and RNA

thymine one of the common nitrogen bases in DNA and RNA

thymine dimers boxy structures resulting from covalent attachment of adjacent thymines in the same DNA strand; formed when DNA is exposed to UV light

tissue typing identification of HLA types

topoisomerase a helicase that untangles DNA by cutting and ligating one or both strands of the double helix

touchdown PCR a modification of the polymerase chain reaction method to enhance amplification of desired targets by manipulation of annealing temperatures

tracking dye in electrophoresis, a dye added to samples being separated to monitor the extent of migration through the matrix

tracrRNA noncoding RNA that forms a ribonucleoprotein complex with crRNA and Cas9 endonuclease in bacteria to target invading DNA

trait a particular characteristic

transcription template-guided synthesis of RNA by RNA polymerase

transcription-mediated amplification TMA; an isothermal amplification method using an RNA intermediate; TMA is also an abbreviation for tissue microarrays

transduction transfer of genetic information from one cell to another through a viral intermediate

transfer in blotting, movement of DNA or protein resolved by electrophoresis from the gel matrix to a membrane

transfer RNA adaptors that link codons in messenger RNA with amino acids

transformation transfer of genetic information among cells without physical association, such that a new phenotype is produced in the recipient cells

translocation breakage and fusion of separate chromosomes

transmembrane spanning across the cell membrane

transmission-based precautions protective equipment and/or clothing used with potential exposure to airborne or contact-transmissible infectious agents

transmission pattern mode of inheritance of a phenotype within a family

transposon a fragment of DNA with the capacity to move from one genetic location to another

trimming addition or deletion of nucleotides at the junctions between V, D, and J segments during immunoglobulin and T-cell receptor gene rearrangements

trinucleotide repeat three 3-bp tandem sequence repeats in DNA

triploid having three of each chromosome

true negative lack of detection of a target marker that is not present in a test sample

true positive accurate detection of a target marker present in a test sample

truncation premature protein termination

***Tth* polymerase** a heat-stable polymerase isolated from *Thermus thermophilus*

tumor-suppressor gene TSG; a gene that dampens proliferation and survival; TSG function is often lost in cancer cells

typing capacity ability of a test to identify the target organism

typing trays plates preloaded with reference antibodies used for study of HLA antigens

tyrosine kinase inhibitor TKI; chemotherapeutic agent that inhibits phosphorylation catalyzed by oncogenes such as *EGFR* and *BCR-ABL*

U

unbalanced in cytogenetics, genetic events that result in the gain or loss of genetic material

uniparental disomy inheritance of chromosomal material from only one parent due to aberrant separation of chromosomes during meiosis

V

vacuum transfer movement of nucleic acid or protein from a gel matrix to a membrane using a vacuum

validation a series of analyte tests using a particular method to establish method characteristics

variable expressivity a range of phenotypes in individuals with the same gene lesion

variable-number tandem repeats VNTR; head-to-tail repeats in DNA with 10- to 150-bp repeat units

V(D)J recombination normal intrachromosomal breaking and joining of DNA in the genes coding for immunoglobulins and T-cell receptors

viral load quantified presence of virus in a specimen

virus a microorganism dependent on host cell machinery for reproduction; also, a sometimes-deleterious electronic signal that can be passed among computers

W

well in gel electrophoresis, an indentation or space molded into the gel to accommodate loading of molecules to be separated

western blot a method used to analyze specific proteins or peptides in a mixture of other proteins

whole chromosome paints fluorescent probes that stain the entire length of a specific chromosome

X

X-linked having genetic components located on the X chromosome

Y

Y-STR short tandem repeats located exclusively on the Y chromosome

Y-STR/paternal lineage test establishment of paternal ancestry using Y chromosome polymorphic short tandem repeat haplotypes

Z

zinc finger specialized Zn-stabilized secondary structure found in proteins that bind DNA

zwitterion a molecule that can be completely positively or negatively charged at a given pH

Index

Numbers followed by *f* represent figures, numbers followed by *t* represent tables, and numbers followed by *b* represent boxes.

A

A site, 49, 49*f*
Absorptivity constants, 92
Acrocentric chromosome, 184, 184*f*
Acrylic shielding, 466, 467*f*
Activator, 63
Acute lymphocytic leukemia, 406*t*
Acute monocytic leukemia, 406*t*
Acute myeloid leukemia, 406*t*
Acute myeloid leukemia/myelodysplastic syndrome, 406*t*
Acute myelomonocytic leukemia, 406*t*
Acute nonlymphocytic leukemia, 406*t*
Acute promyelocytic leukemia, 406*t*
Acute T-lymphocytic leukemia, 406*t*
Adaptors, 16
Adenine, 3
Adenosine, 4
Adler, Julius, 12*b*
Affinity maturation, 391*b*
AFLP analysis, 334*f*
Agarobiose, 99*f*
Agarose, 98, 99*f*, 99*t*
Agarose gel electrophoresis, 91*f*
Agarose gels, 99–101
Alanine tRNA, 34*f*
Algorithm, 249*t*
Alignment, 249*t*
Alignments of HLA polymorphic regions, 431
Alkaline phosphatase, 132*b*, 201, 202*b*, 202*f*, 227*b*
Allele dropout, 242
Allele peak locations, 269*f*
Alleles
  defined, 420
  HLA nomenclature, 421–422, 425
  HLA polymorphisms, 420
  identification serologically and by DNA sequence, 426*t*
  locus versions, 263
  null allele, 425
Allele-specific oligomer hybridization (ASO), 209
  dot blot, 209
  reverse dot blot, 209
Allele-specific primer amplification, 214*f*
Allelic discrimination with fluorogenic probes, 214–215, 215*f*
Allelic exclusion, 388
Allelic frequencies, paternity testing, 275–277
  combined paternity index, 276
  paternity index, 275–276
  prior odds, 276–277, 277*t*
  probability of paternity, 276–277
Allelic ladders, 268
Alloantibodies, 428
Allogeneic transplant, 283
Allografts, 418
Alpha helix, protein, 42, 42*f*
Alpha phosphate, 29, 59
Alpha satellite, 183
Alternative nucleic acids, 126*b*
Alternative splicing, 33, 121–122
Alu elements, 261
Ambiguity, 425
Amelogenin locus, 270, 270*f*
Amino acid, 38–43, 39*f*
Amino acid biosynthetic groups, 40*t*
Amino acid charging, 46–47
Amino acid classification, 40*t*
Amino acyl tRNA synthetases, 47*f*
Amino-terminal, 38
Aminoacyl tRNA synthetases, 46
Ammonium persulfate (APS), 102
Amplicons, 143, 147*f*
Amplification
  control, 153, 305, 306*f*, 457
  probe, 168–171
  program, 144
  signal, 172–174
  target, 143–168

Amplified fragment length polymorphism, 332–333
Analyte measurement range, 454*t*
Analyte-specific reagents (ASR), 464
Analytic accuracy, 454, 454*t*
Analytic sensitivity, 454*t*
Analytic specificity, 454*t*
Analytical targets, molecular testing, 372
  acute lymphocytic leukemia, 406*t*
  acute monocytic leukemia, 406*t*
  acute myeloid leukemia, 406*t*
  acute myeloid leukemia/myelodysplastic syndrome, 406*t*
  acute myelomonocytic leukemia, 406*t*
  acute nonlymphocytic leukemia, 406*t*
  acute promyelocytic leukemia, 406*t*
  acute T-lymphocytic leukemia, 406*t*
  anaplastic lymphoma receptor tyrosine kinase (*ALK*) proto-oncogene, 2p23.1, 382
  aneuploidy, 372
  ataxia telangiectasia mutated gene, *ATM* (11q22), 379–380
  B-cell leukemia, 406*t*
  breast cancer 1 gene, *BRCA1* (17q21), and breast cancer 2 gene, *BRCA2* (13q12), 380
  Burkitt lymphoma, 406*t*
  chronic lymphocytic leukemia, 406*t*
  chronic myelogenous leukemia, 406*t*
  clonality detection, 391
    banding patterns, 396–397
    complementarity determining regions, 393
    consensus primer, 393
    framework regions, 393
    hematological malignancies, gene mutations, 404–405
    immunoglobulin heavy chain gene rearrangements, 391–394
    immunoglobulin light chain gene rearrangements, 394
    mutations, hematological malignancies, 397–405
    polyclonal lymphocytes, 391
    T-cell receptor gene rearrangements, 394–396
  diffuse large B-cell lymphoma, 406*t*
  epidermal growth factor receptor, *EGFR* (7p12), 373–375
  Ewing sarcoma, *EWS* (22q12), 376–377, 377*t*, 378*f*
  follicular lymphoma, 406*t*
  gene/chromosomal mutations, solid tumors, 372–387
  hairy cell leukemia, 406*t*
  Harvey rat sarcoma viral oncogene homolog, *H-ras* (11p15), 375–376, 375*f*, 375*t*
  hematological malignancies, molecular analysis, 387–406
  hematological malignancies, molecular targets
    allelic exclusion, 388
    gene rearrangements in leukemia and lymphoma, 387–397
    germline, 387–388
    immunoglobulin heavy-chain gene re-arrangement, B cells, 387–388
    V(D)J recombination, 387, 387*f*
  heterozygosity loss, 385, 386*f*
  human epidermal growth factor receptor 2, *HER2/neu/erb-b2 1* (17q21.1), 372–373
  Kirsten rat sarcoma viral oncogene homolog, K-ras (12p 12), 375–376, 375*f*, 375*t*, 376*f*
  leukemia and lymphoma molecular analysis
    affinity maturation, 391*b*
    immunoglobulin light-chain gene re-arrangement in B cells, 389–390
    kappa deleting element, 389
    somatic hypermutation, 388
    T-cell receptor gene rearrangement, 390–391
    trimming, 391
  leukemias, chromosomal abnormalities, 406*t*
  liquid biopsy, 386–387
  lymphomas, chromosomal abnormalities, 406*t*
  mantle cell lymphoma, 406*t*
  microsatellite instability, 382, 384–385, 385*f*
  molecular abnormalities, 382
  mucosa-associated lymphoid tissue lymphoma, 406*t*
  multiple myeloma, 406*t*
  myeloproliferative/myeloblastic disease, 406*t*
  myeloproliferative/myelodysplastic disease, 406*t*
  neuroblastoma ras, *N-ras* (1p13), 375–376, 375*f*, 375*t*
  paired box-forkhead in rhabdomyosarcoma, *PAX3-FKHR*, *PAX7-FKHR*, t(1;13), t(2;13), 378
  polycythemia vera, 406*t*
  pre-B acute lymphoblastic leukemia, 406*t*
  proto-oncogenes, 381
  rearranged during transfection proto-oncogene (10q11), 381
  rearranged during transfection (*RET*) proto-oncogene (10q11), 381
  sequencing panels, 405
  synovial sarcoma translocation, chromosome 18-synovial sarcoma breakpoint 1 and 2, *SYT-SSX1*, *SYT-SSX2* t(X;18) (p11.2;q11.2), 377–378
  T-cell lymphoma, 406*t*
  T-chronic lymphocytic leukemia, 406*t*
  tissue-specific, 372
  tumor protein 53, *TP53* (17p13), 378–379
  tumor-specific, 372
  V-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog, *Kit*, *c-KIT* (4q12), 382
  V-myc avian myelocytomatosis viral-related oncogene, neuroblastoma-derived, 381
  Von Hippel-Lindau gene, *VHL* (3p26), 380–381
  V-Ros Avian UR2 Sarcoma Virus Oncogene Homolog 1 (*ROS1*) Proto-Oncogene (6q22.1), 381
  Waldenström macroglobinemia, 406*t*
Anaplastic lymphoma receptor tyrosine kinase *(ALK)* proto-oncogene, 2p23.1, 382
Anderson reference, 291
Aneuploid cells, 181
Aneuploidy, 181, 187*f*, 188*f*, 345, 372
Anion, 123
Annealing, 145, 146*f*, 147*b*
Annotation, 246, 249*t*
Anode, 98, 120
Anticodon, 34, 45–47
Antigen detection, 202*f*
Antigen retrieval, 202
Antihuman antibodies, 428
Antimicrobial action sites, 325*f*
Antimer, 70
Antimicrobial agents, 324–329
  bacteriocidal, 324
  bacteriostatic, 324
  fungicidal, 324

  fungistatic, 324
  mode of action, 325*t*
  molecular detection of resistance, 327–329
  resistance to, 302, 325–327
Antimicrobial resistance in *M. tuberculosis*, 329
Antiparallel, 9
Antisense strand, 59
Apoptosis, 183, 183*f*, 371
Apoptotic DNA, 183*f*
APS. *See* Ammonium persulfate (APS)
Apurinic site, 119*f*
Arbitrarily primed PCR, 167–168, 169*f*, 332
Array CGH, 363
Array comparative genome hybridization technology, 363
Array technology, 212–214
  bead array, 213
  high-density oligonucleotide arrays, 212
  redundant tiling, 212
  standard tiling, 212, 212*f*
Array-based hybridization, 133–138
  dot blot, 133–134
  genomic array technology, 134–138
    macroarrays in, 134
    microarrays in, 134–138, 135*f*
    microelectronic arrays, 135–138
  slot blot, 133–134
Asexual reproduction, recombination in, 21–25
  conjugation, 21–23
  transduction, 23
  transformation, 23–25
ASO. *See* Allele-specific oligomer hybridization (ASO)
ASR. *See* Analyte-specific reagents (ASR)
Assign, 235*t*
Ataxia telangiectasia, 379
Ataxia telangiectasia mutated gene, *ATM* (11q22), 379–380
Attenuation, 63, 63*f*
Autologous bone marrow transplant, 283, 283*f*
Automated DNA isolation, 85*f*
Automated fluorescent sequencing, 229–235, 230*f*
  cleaning the sequencing ladder, 232
  dye primer, 230, 231*f*
  dye terminator, 230, 231*f*
  electrophoresis, 231–232
  sequence interpretation, 232, 233*f*–234*f*
Autophagy, 349
Autoradiograph, 469
Autosomal STRs, 278
Autosomal-dominant transmission, 347, 347*f*
Autosomal-recessive mutations, 348, 348*f*
Avuncular testing, 277–278

B

Bacteria, 309–313
  DNA isolation, 80
  respiratory tract pathogens, 309–311
  urogenital tract pathogens, 311–313
Bacteriocidal, 324
Bacteriocins, 26
Bacteriophage, 14, 23
Bacteriophage T4, 24*f*
Bacteriostatic, 324
Balanced chromosome, 187
Balanced polymorphism, 181
Balanced reciprocal translocation, 189*f*
Band, 90–91, 99
Banding patterns, 396–397
Bar coding, 240, 241*f*
Barr body, 345*b*
Basal transcription complex, 28
Base calling, 232
Base of RNA, 27
Base pairing, 105
Base pairs, 70
Basic Local Alignment Search Tool (BLAST), 235*t*, 248
Basic PCR procedure, 143–147
B-cell leukemia, 406*t*
bDNA amplification, alkaline phosphatase, 172–173, 172*f*
Bead array technology, 138, 213, 213*f*
Bead arrays, serum antibody detection, 429*f*
Benzer, Seymour, 23, 44*b*
Bessman, Maurice J., 12*b*
Beta-lactam antibiotic resistance, 327–328
Beta-lactamase, reaction, 328*f*
Beta-pleated sheets, 42
Bilateral symmetry, 16
Bin, 273
Binding sites, 8
Binning, 273, 273*b*
Biochemical methods, 201–207
  alkaline phosphatase, 201, 202*b*, 202*f*
  enzyme immunoassays, 201–202, 202*b*
  gas chromatography, 204–205
  high-performance liquid chromatography, 204
  immunoassays, 201–202
  immunosorbent assays, 202*b*
  mass spectrometry, 205–207
Biochemistry, nucleic acid, 2–37
Biohazard stickers, 447*f*
Bioinformatics, 248–250
  algorithm, 249*t*
  alignment, 249*t*
  annotation, 249*t*
  computational biology, 248
  conservation, 249*t*
  domain, 249*t*
  gap, 249*t*
  GenBank, 249*t*
  homology, 249*t*
  identity, 249*t*
  interface, 249*t*
  local alignment, 249*t*
  motif, 249*t*
  multiple sequence alignment, 249*t*
  optimal alignment, 249*t*
  orthology, 249*t*
  paralogy, 249*t*
  PubMed, 249*t*
  query, 249*t*
  in silico, 248
  similarity, 249*t*

Bioinformatics *(cont'd)*
  SwissProt, 249*t*
  terminology, 249*t*
Biology, molecular, 77–257
  chromosomal structure and mutations, 179–198
  DNA sequencing, 223–257
  gene mutations, 199–222
  nucleic acid amplification, 142–178
  nucleic acid and protein analysis and characterization, 112–141
  nucleic acid extraction methods, 78–96
  nucleic acid resolution and detection, 97–111
Biotin, 126, 127*f*
Bisulfite DNA sequencing, 236–237
BK/JC viruses, 323
BLAST, 235*t*, 248
Block thermal cyclers, 460, 461*f*
Blocking, 123
Blunt ends, 16
Bone marrow aspirate. *See* Nucleated cells in suspension
Bone marrow engraftment testing, DNA polymorphisms, 283–289
  allogeneic transplant, 283
  autologous transplant, 283
  chimera, 284
  fragment analysis, 286
  full chimerism, 285
  graft-versus-host disease, 284
  graft-versus-tumor, 284
  hematopoietic stem cells, 283
  informative locus, 285
  linkage analysis, 282–283, 282*f*
  matched unrelated donor, 284
  mixed chimerism, 285
  mosaic, 284*b*
  myeloablative, 284, 287
  noninformative loci, 285
  post-transplant engraftment testing, 287–289, 287*f*
  pretransplant STR testing, 285–287
  split chimerism, 286*b*
  stutter, 286–287
Bordetella pertussis, 309
Bover, K., 418*b*
Box-forkhead in rhabdomyosarcoma, 378
Branched DNA amplification (bDNA), alkaline phosphatase, 172–173, 172*f*
Break-apart probes, 192, 192*f*
Breast cancer 1 gene, *BRCA1* (17q21), 380
Breast cancer 2 gene, *BRCA2* (13q12), 380
Brenner, Sydney, 43
Bridge PCR, 167, 168*f*. *See also* Surface amplification
Buffer, agarose gels, 99
Buffer systems, 104–106
  additives, 105–106
    denaturing agents, 105
    ph drift, 106
  conjugate, 104
Burkitt lymphoma, 406*t*

C

C banding, 185*b*
Calibration curve, 457
Calibration verification, 457
Calibrations, 456–457, 463
Cambridge reference sequence, 291
Cameras, gel photography, 462*f*
Cameras and detectors, 462
Cancer, defined, 370
Cancer molecular basis, 371–372
  apoptosis, 371
  cell division cycle, 371, 371*f*
  gain-of-function, 371
  loss-of-function, 371–372
  oncogenes, 371
  tumor-suppressor genes, 371
Cap, 31, 32*b*
Capillary electrophoresis, 37, 102–104, 103*f*
Capillary gel electrophoresis, 269*f*, 384, 385*f*
Capillary transfer, 120, 120*f*
Capping, 31
Carbon position numbering, nucleotide, 6*f*
Carboxy-terminal, 38
Carcinomas, 370
Cathode, 99, 120
CDC assay example, 428*t*
CDC expression, 428*t*
cDNA, 157, 165*f*
CDR. *See* Complementarity determining region (CDR)
Cell division cycle, 371, 371*f*
CEN probes, 192, 194*f*
Centrifuge speeds, 461*f*
Centrifuges, 461
Centromere, 183, 184*f*
Centromeric probe, 192, 194*f*
Certified chamber thermometers, 460*f*
CGH. *See* Comparative genome hybridization (CGH)
Chaperones, 49, 50*f*
Charging reaction, 46
Chase, Martha, 23
Chemical cleavage, 215–216
Chemical safety, 465–466
Chemical sequencing, 224–225, 224*f*
Chimera, 284
Chimerism, 345
Chlamydia, 311
Chlamydia trachomatis, 311
Chlamydophila pneumoniae, 311
Chromatin, 18, 65
Chromosomal abnormalities, 345–346
  aneuploidy, 345
  chimerism, 345
  marker chromosomes, 346
  mutations, 347*t*
  polyploidy, 345
  triploid, 346
Chromosomal mutations, 253*t*
Chromosomal structure, 181–186
  aneuploid cells, 181
  aneuploidy, 187*f*, 188*f*
  apoptotic DNA, 183*f*
  balanced polymorphism, 181
  balanced reciprocal translocation, 189*f*

  break-apart probes, 192*f*
  centromere, 183, 184*f*
  centromeric probe, 194*f*
  chromosomal compaction, histones, 181–183, 182*f*
  chromosome morphology, 183–184
  chromosome mutations, 181, 190*f*
  classification by size and centromere position, 185*t*
  diploid, 180, 181
  euploid, 181
  FISH analysis, 192*f*
  G-band patterns, 185*f*
  gene mutations, 181
  genome, chromosomal mutations, 186–195
    fluorescence in situ hybridization, 190–195
    karyotyping, 186–190
  genome mutations, 181
  genotype, 180
  haploid, 180
  human genome, 180
  karyotype, showing balanced reciprocal translocation, 189*f*
  male karyotype, 187*f*
  metacentric chromosomes, 184*f*
  metaphase chromosomes, 182*f*
  mutation, 180–181
  phenotype, 180
  polymorphism, 180
  Robertsonian translocation, 189*f*
  sequence of nucleotides, 180
  sizes of human chromosomes, 180*t*
  staining pattern, chromosomes, 185*f*
  telomeric probes, 193*f*
  trisomy, 181
  variant, 180
  visualizing chromosomes, 184–186
Chromosome, 43
Chromosome mutations, 181, 190*f*
Chromosome spread, 186
Chronic lymphocytic leukemia, 406*t*
Chronic myelogenous leukemia, 406*t*
Cis factors, 60, 61*f*
Clades, 314
Class switching, 388
Class-switch recombination, 389*b*
Cleavage-based amplification, 173–174, 174*f*
Cleavase assay, 216, 218*f*
Cleavase single-color assay, 218*f*
CLIA. *See* Clinical Laboratory Improvement Amendments (CLIA)
Clinical Laboratory Improvement Amendments (CLIA), 447, 470
Clinical sensitivity, 454, 454*t*
Clinical specificity, 454, 454*t*
Clonality
  detection. *See* Clonality detection
  molecular analysis of immunoglobulin heavy-chain gene clonality, 391–394, 392*f*, 393*f*, 394*f*
  monoclonal, 391
  polyclonal, 391
Clonality detection, 391
  banding patterns, 396–397
  complementarity determining regions, 393
  consensus primer, 393
  framework regions, 393
  hematological malignancies, gene mutations, 404–405
  immunoglobulin heavy chain gene rearrangements, 391–394
  immunoglobulin light chain gene rearrangements, 394
  mutations, hematological malignancies, 397–405
  polyclonal lymphocytes, 391
  T-cell receptor gene rearrangements, 394–396
  translocations, hematological malignancies
    RNA integrity control, 401
    t(8;14)(q24;q11), 400, 400*f*
    t(9;22)(q34;q11), 401–403, 401*f*, 402*f*, 403*f*
    t(11;14)(q13;q32), 398–399, 400*f*
    t(14;18)(q32;q21), 397–398, 397*f*, 398*f*, 399*f*
    t(15;17)(q22;q11.2-q12), 403–404, 404*f*
Clonotype, 396*b*
Clustered, regularly interspaced, short palindromic repeat (CRISPR), 17*b*
CMV. *See* Cytomegalovirus (CMV)
Coa typing, 335
Coagulase-negative *Staphylococcus*, 331*f*
CODIS. *See* Combined DNA indexing system (CODIS)
Codominant offspring, 347
Codons, 45, 45*f*
Coefficient of variance, 457, 461
Colicinogenic factors, 26
Collection tubes
  for molecular testing, 448–450
  for nucleic acid testing, 449*f*, 449*t*
Comb, 106
Combined DNA indexing system (CODIS), 272*b*
Combined paternity index (CPI), 276
Combined sibling index, 277
Combining typing results, 436, 436*t*
Compaction, histones, 181–183, 182*f*
Comparative genome hybridization (CGH), 136–137, 195, 195*f*, 196*f*
Comparative genome hybridization technology, 196
Comparative genomic arrays, 166
Complement, 44*b*
Complementarity determining region (CDR), 393
Complementary DNA, 31, 157
Complementary nucleotide, 7
Complement-dependent cytotoxicity, 427
Complete dominance, 347
Complete penetrance, 349
Components of PCR, 144*f*, 144*t*, 147–151
Computational biology, 248
Concatamers, 19
Conformer, 207
Congenital diseases, 345
Conjugate, 104
Conjugated proteins, 43
Conjugating cells, 22*f*
Conjugation, 21–23
Consensus, 42*b*
Consensus primer, 393
Consensus sequences, 248
Conservation, 249*t*
Conservative, 201
Conservative substitutions, 200
Constitutive RNA, 30, 60

Contact precautions, 450
Contaminants, absorbance, 93*t*
Contamination, 150*b*
Contamination control, 152–153, 305
Continuous allele system, 266
Controls
    PCR, 152–156
    PCR contamination, 153–154
    test performance, 457–458
Coordination of HLA test methods, 437
Corepressor, 63
Corey, Robert, 42
C_0 t value, 128
C_0 t½, 128
Coverage, 247
CpG islands, 66, 68*f*
CPI. *See* Combined paternity index (CPI)
CREG. *See* Cross-reactive epitope group (CREG)
Crick, Francis, 43, 44*b*, 45
CRISPR enzyme systems, 115–116, 116*f*
crRNA, 115
Criteria for accepting specimens, 447–448, 448*f*
Crossing over, 20, 20*f*
Crossmatching
    antibodies, 426, 427*f*
    serological analysis, 429, 430*f*
Cross-reactive epitope group (CREG), 428
Crude DNA extraction, 86*f*
Cryostat, 202
Cryotubes, 451, 453*f*
C_T, 160
Curve analysis, homozygous mutant, 209*f*
Cut points, 458
Cut-off values, 458
Cycle sequencing, 229
Cycles, 144
Cycling probe, 174
Cycling probe fluorescence, 174*f*
Cystic fibrosis, 353–354, 354*f*
Cystic fibrosis transmembrane conductance, 353, 354*f*
Cytidine, 4
Cytochrome P-450 enzymes, 354–355, 354*f*, 355*f*
Cytomegalovirus (CMV), 320–321
Cytopathic effect, 313–314
Cytosine, 3
Cytosine residues, 238*f*
Cytotoxicity, cells stained for, 428*f*

D

DAPI. *See* Diamidino-2-phenylindole (DAPI)
ddNTP. *See* Dideoxynucleotide (ddNTP)
Deaza dGTP, 149*b*
Deletion, 187
Denaturation
    amplification program, 144
    DNA, 119
    DNA target, 145*f*
Denaturing agents, 8*b*, 105
Denaturing gels, 123
Deoxynucleotide triphosphates (dNTP), 143, 146*f*
Deoxyribonuclease, 25
Deoxyribonuclease I, 18
Deoxyribonucleic acid. *See* DNA
Deoxyribonucleotide bases, 149–150
Deoxyribose, 3
Depurination, 118, 119*f*
Derivative chromosome, 188
Detection
    microorganisms, 301–343
        AFLP analysis, 334*f*
        amplification controls, 306*f*
        antimicrobial action sites, 325*f*
        antimicrobial agents, 302, 324–329
        bacteria, 309–313
        beta-lactamase, reaction, 328*f*
        coagulase-negative *Staphylococcus*, 331*f*
        epidemiological typing methods, 330*t*
        fluorescent AFLP analysis, 335*f*
        fungi, 323–324
        genes conferring resistance to antimicrobial agents, 326*t*
        genital tract organisms, 312*t*
        genomes, human viruses, 314*t*
        HIV viral loads, nucleic acid amplification methods, 318*t*
        molecular detection, microorganism, 308–324
        molecular epidemiology, 329–337
        nucleic acid amplification tests, viruses, 315*t*–317*t*
        nucleic acid test, target sequences, 307*f*
        parasites, 324
        penicillin, 328*f*
        penicillin structure, 328*f*
        PFGE pattern interpretation, 332*t*
        quality control, 305–306
        RAPD gel results, 333*f*
        resistance mechanisms, 326*t*
        respiratory tract organisms, 310*t*
        reverse transcriptase, 302
        ribosomal RNA genes, 336*f*
        sample preparation, 304–305
        sequence targets, 307
        specimen collection, 302–304
        specimen transport systems, 303*t*
        Swab Extraction Tube System, 303*f*
        vancomycin structure, 329*f*
        vancomycin-resistant *S. aureus*, 327*f*
        viral load measurement, test performance, 318*t*
        viruses, 313–323
    nucleic acid, 97–111
        agarobiose, 99*f*
        agarose, 99*t*
        buffer systems, 104–106
        capillary electrophoresis, 103*f*
        detection systems, 109–111
        double-stranded DNA, 100*f*
        double-stranded DNA fragments, 102*f*
        dye comigration, 109*t*
        electrophoresis, 98
        electrophoresis equipment, 106–109
        field inversion gel electrophoresis, 100*f*
        gel loading, 108–109
        gel systems, 99–102

    horizontal gel electrophoresis, 98*f*
    horizontal submarine gel, 106*f*
    peristaltic pump, 106*f*
    polyacrylamide, 101*f*
    polyacrylamide concentration, 103*t*
    polyacrylamide electrophoresis combs, 107*f*
    vertical gel apparatus, 108*f*
Detection limit, 454*t*, 455
Detection of clonality. *See* Clonality detection
Detection of gene mutations, 201–218
  biochemical methods, 201–207
  chemical cleavage, 215–216
  Cleavase assay, 216, 218*f*
  enzymatic and chemical cleavage methods, 215–217
  heteroduplex analysis with single-strand specific nucleases, 216
  hybridization-based methods
    allele-specific oligomer hybridization, 209
    array technology, 212–213
    heteroduplex analysis, 211–212
    melt-curve analysis, 209–211
    single-strand conformation polymorphism, 207–209
  methods, 217–218
  nonisotopic RNase cleavage assay, 216, 217*f*
  nucleic acid analyses, 207–218
  restriction fragment length polymorphisms, 215–216, 215*f*, 216*f*
  sequencing-based methods, 213–215
    allelic discrimination with fluorogenic probes, 214–215, 215*f*
    sequence-specific PCR, 213–214
Detection systems, 109–111, 129–132
  fluorescent dyes, 109–110
  microorganisms, genes conferring resistance to antimicrobial agents, 326*t*
  nonradioactive detection, 130, 131*f*
  nucleic acid-specific dyes, 110
  probe labeled with radioactive phosphorous atoms, 130, 130*f*
  signal-to-noise ratio, 132
  silver stain, 110–111
dGMP. *See* Nucleotides deoxyguanosine monophosphate (dGMP)
Diamidino-2-phenylindole (DAPI), 186
Dideoxy chain termination (Sanger) sequencing, 225–229
  alkaline phosphatase, 227*b*
  ddNTP derivatives added, 226*f*
  dideoxynucleotide, 226, 227*f*
  internal labeling, 226
  M13, 226*b*
  M13 universal primer, 226*b*
  manual, 225*f*
  sequencing ladder, 228, 229*f*, 230
Dideoxynucleotide (ddNTP), 226, 226*f*, 227*f*
Diffuse large B-cell lymphoma, 406*t*
Digoxygenin, 126
Dioxetane substrates, light emitted, 130*f*
Diploid, 180
Diploid organisms, 181
Direct sequencing, 224–235
  automated fluorescent sequencing, 229–235, 230*f*
    cleaning the sequencing ladder, 232
    dye primer, 230, 231*f*
    dye terminator, 230, 231*f*
    electrophoresis, 231–232
    sequence interpretation, 232, 233*f*–234*f*
  manual sequencing, 224–229
    chemical sequencing, 224–225, 224*f*
    dideoxy chain termination (Sanger) sequencing, 225–229, 225*f*, 226*f*
    Maxam-Gilbert sequencing, 224–225
Discrete allele systems, 268
Discriminatory capacity, 279
Discriminatory power, 337
Dispersive, 10*b*
DNA, 2–20
  adaptors, 16
  adenine, 3
  adenosine, 4
  antiparallel, 9
  bacteriocins, 26
  bacteriophage, 14, 23
  bacteriophage T4, 24*f*
  bilateral symmetry, 16
  blunt ends, 16
  carbon position numbering, nucleotide, 6*f*
  colicinogenic factors, 26
  complementary, 16
  complementary nucleotide, 7
  concatamers, 19
  conjugating cells, 22*f*
  CpG islands, 66
  crossing over, 20, 20*f*
  cytidine, 4
  cytosine, 3
  demethylation, 69
  denaturing agents, 8*b*
  deoxyribonuclease, 25
  deoxyribonuclease I, 18
  deoxyribose, 3
  dispersive, 10*b*
  DNA polymerase, 12*f*, 13*f*
  DNA replication, 8*f*
  DNA synthesis, 9*f*
  double helix, 4*b*, 5*f*, 7*b*, 9*b*
  double stranded, 8
  double-strand break, 16, 19
  endonucleases, 14
  enzymes metabolizing DNA, 14–19
  episomal nature of the resistance factor, 25–26
  episomes, 26
  exonuclease, 12
  exonuclease I, 17
  exonuclease II, 18
  exonuclease III, 17
  exonuclease VII, 17
  fertility factor, 21, 22*f*
  fidelity, 14
  gamete, 14
  gap-filling DNA polymerase, 11*b*
  guanine, 3
  guanosine, 4
  gyrases, 19
  hemimethylated, 19

DNA *(cont'd)*
  heterosis, 19
  Hfr, 23
  histone code, 65
  holoenzyme, 11–12
  homologous sequences, 9*f*
  hybridization, 9
  hydrogen bonds, 7
  imprinting, 66
  intercalating agents, 8*b*
  Klenow fragment, 12
  lagging strand, 10
  leading strand, 10
  ligase, 13
  major groove, 8*b*
  meiosis, 20
  Mendel, 19*b*–20*b*
  methylation, 66–68
  micrococcal nuclease, 18
  migration, 25
  minor groove, 8*b*
  mismatches, 7
  monomeric, 14
  mung bean nuclease, 18
  mutations, 12
  nick, 13
  nick translation, 13, 13*f*
  nitrogen base, 3, 7
  nuclease BaI31, 18
  nuclein, 3
  nucleoside, 4, 5*f*
  nucleotide deoxyguanosine 5′ phosphate, 5*f*
  nucleotides, 3
  nucleotides deoxyguanosine monophosphate, 6*f*
  Okazaki fragments, 10
  oligonucleotides, 18
  overhangs, 16
  palindromic, 16
  phosphodiester bond, 3
  plasmids, 23, 25–26, 26*f*
  polymerase classified by sequence homology, 11*t*
  polymerases, sequence homology classification, 11*t*
  primase, 10
  processivity, 14
  proofread, 13
  protease, 24–25
  purines, 4
  pyrimidine dimers, 14*b*
  pyrimidines, 4
  pyrophosphate exchange, 12
  pyrophosphorolysis, 12
  R factors, 26
  radioactive protein, 24*f*
  recBC nuclease, 18
  recognition sites, 8*b*
  recombinant, 19, 21
  recombination, 19, 21*f*, 22*f*
  recombination in asexual reproduction, 21–25
  recombined chromosomes, 22*f*
  replication, 9–14
  replication complex, 11*b*
  replication fork, 10
  replisome, 11*b*
  restriction enzymes, 14–17, 15*t*, 16*f*
  restriction modification, 7*b*
  ribonuclease, 24–25
  RNase H, 11*b*
  S1 nuclease, 18
  satellite plasmid, 26
  semiconservative, 9
  sexually reproducing organisms, recombination in, 19–21
  simultaneous replication, 10*f*
  single-strand break, 19
  sticky ends, 16
  structure, 4–9
  substrate specificity, 14
  sugar-phosphate backbone, 8
  supercoiled, 19
  supercoiled plasmids, 26, 26*f*
  template, 9
  terminal transferase, 14
  thymidine, 4
  thymine, 3, 14*b*
  topoisomerases, 19
  traits, 3
  transduction, 23
  transformation, 23–25
  transforming factor, 24*f*, 25
DNA extraction/storage cards, 86
DNA fingerprinting, 265–266, 266*b*
DNA isolation, 79–87
  bacteria and fungi, 80
  inorganic isolation methods, 83
  isolation of mitochondrial DNA, 86
  lipoproteins, 81
  nucleated cells in suspension, 80
  organic isolation methods, 81–83
  plasma, 80–81
  preparing sample, 79–81
  proteolytic lysis, 85–86
  salting out, 83
  solid-phase isolation, 84–85, 84*f*
  specimen sources, 79*t*
  tissue samples, 81
  viruses, 80
DNA metabolizing enzymes, 17–19
DNA polymerase, 12*f*, 13*f*, 146*f*, 150–151
DNA polymorphisms, 260–300
  allele peak locations, 269*f*
  amelogenin locus, 270*f*
  autologous bone marrow transplant, 283*f*
  bone marrow engraftment testing, 283–289
    allogeneic transplant, 283
    autologous transplant, 283
    chimera, 284
    fragment analysis, 286
    full chimerism, 285
    graft-versus-host disease, 284
    graft-versus-tumor, 284
    hematopoietic stem cells, 283

    - informative locus, 285
    - linkage analysis, 282–283, 282*f*
    - matched unrelated donor, 284
    - mixed chimerism, 285
    - mosaic, 284*b*
    - myeloablative, 284, 287
    - noninformative loci, 285
    - post-transplant engraftment testing, 287–289, 287*f*
    - pretransplant STR testing, 285–287
    - split chimerism, 286*b*
    - stutter, 286–287
- capillary gel electrophoresis, 269*f*
- electropherogram, 276*f*, 278*f*
- epigenetic profiles, 294
- HLA-DRB1 gene, 421*f*
- linkage analysis, 282–283, 282*f*
- microvariant allele, 272*f*
- mitochondrial DNA polymorphisms, 291–293
- mitochondrial genome, 291*f*
- PAGE analysis, 285*f*
- paternity tests, 276*t*
- polymorphisms, 262*t*
- protein-based identification, 293–294
- quality assurance for surgical sections using STR, 289, 289*f*
- RFLP inheritance, 264*f*
- RFLP typing, 262–266
    - alleles, 263
    - genetic mapping with RFLPs, 264
    - heterozygous, 263
    - homozygous, 263
    - human identification, RFLP, 265–266
    - locus, 263
    - parentage testing, RFLP, 264–265, 265*f*
- short tandem repeat TH01, 267*f*
- short tandem repeats, 268*f*
- single nucleotide polymorphisms, 289–291
    - Human Haplotype Mapping project, 290–291, 290*f*, 291*b*
    - Southern blot, 262*t*
    - types, 290*t*
- STR locus information, 271*t*
- STR typing by PCR, 266–278
    - allelic ladders, 268
    - discrete allele systems, 268
    - gender identification, 270, 270*f*
    - internal size standards, 268
    - microsatellites, 267
    - microvariants, 267
    - mini-STR, 268
    - short tandem repeats, 267
    - STR analysis, 268–278
    - STR nomenclature, 270
    - test results analysis, 270, 272, 272*t*
- tandem repeat, 265*f*
- types of polymorphisms, 261
- Y-STR, 278–282
    - autosomal STRs, 278
    - genotypes, 281*t*
    - locus information, 280*t*–281*t*
    - matching with Y-STRs, 279, 281–282
    - paternal lineage test, 279

DNA probes, 123–124
DNA replication, 8*f*
DNA sequencing, 223–257
- Assign, 235*t*
- bioinformatics, 248–250
    - algorithm, 249*t*
    - alignment, 249*t*
    - annotation, 249*t*
    - computational biology, 248
    - conservation, 249*t*
    - domain, 249*t*
    - gap, 249*t*
    - GenBank, 249*t*
    - homology, 249*t*
    - identity, 249*t*
    - interface, 249*t*
    - local alignment, 249*t*
    - motif, 249*t*
    - multiple sequence alignment, 249*t*
    - optimal alignment, 249*t*
    - orthology, 249*t*
    - paralogy, 249*t*
    - PubMed, 249*t*
    - query, 249*t*
    - in silico, 248
    - similarity, 249*t*
    - SwissProt, 249*t*
    - terminology, 249*t*
- bioinformatics terminology, 249*t*
- bisulfite DNA sequencing, 236–237
- BLAST, 235*t*
- consensus sequences, 248
- cytosine residues, 238*f*
- dideoxynucleotide, 226*f*
- direct sequencing, 224–235
    - automated fluorescent sequencing, 229–235, 230*f*
    - manual sequencing, 224–229
- electropherogram, 233*f*
- Factura, 235*t*
- FASTA, 235*t*
- fluorescent sequencing chemistries, 231*f*
- GRAIL, 235*t*
- Human Genome Project, 250–254, 250*t*, 251*t*, 253*t*
- IUB universal nomenclature, 248
    - Basic Alignment Search Tool, 248
    - hierarchal shotgun sequencing, 251, 252*f*
    - for mixed bases, 250*t*
- manual dideoxy sequencing, 225*f*
- Matchmaker, 235*t*
- Maxam-Gilbert sequencing, 224*t*, 225*f*
- next-generation sequencing, 238–248, 239*b*, 239*f*, 456
- Phrap, 235*t*
- Phred, 235*t*
- Polyphred, 235*t*
- pyrogram, 236
- pyrosequencing, 235–236, 236*f*
- RNA sequencing, 237–238
- SeqScape, 235*t*
- sequencing ladder, 230

DNA sequencing *(cont'd)*
  software programs, sequence data, 235*t*
  TIGR Assembler, 235*t*
DNA synthesis, 9*f*
DNA template, 149
DNA-based tissue typing, 417–445, 418
  additional recognition factors, 437–438
  allografts, 418
  bead arrays, serum antibody detection, 429*f*
  CDC assay example, 428*t*
  CDC expression, 428*t*
  crossmatching, antibodies, 427*f*
  cytotoxicity, cells stained for, 428*f*
  DNA polymorphisms, 421*f*
  graft-versus-host disease, 418
  HLA polymorphisms, 420–425
    allele, 420, 421–422
    haplotype, 420, 421*f*
    HLA nomenclature, 420, 422–425
    HLA type, 420
  KIR and HLA gene interactions, 438*t*
  KIR gene cluster, 438*f*
  laboratory testing summary, 439–440
  major histocompatibility complex, 418, 419*t*
  MHC disease association, 438–439
  MHC locus, 418–420
  molecular analysis of MHC, 425–437
    combining typing results, 436, 436*t*
    coordination of HLA test methods, 437
    crossmatching, 427
    DNA-based typing, 430–436
    HLA test discrepancies, 436–437
    serological analysis, 425–437
  polypeptides, 420*f*
  reverse dot-blot SSOP, 431–432, 432*f*
  SSOP assay, 431*f*
  transplant evaluation, 439*t*
DNA-based typing, 430–436
  alignments of HLA polymorphic regions, 431
  next-generation sequencing libraries, 239*f*, 435*f*
  other methods, 435–436
  sequence-based typing, 433–435, 433*f*, 434*f*
  sequence-specific oligonucleotide probe hybridization, 431–432
  sequence-specific PCR, 432–433, 432*f*, 433*f*
DNAse I, 18
dNTP. *See* Deoxynucleotide triphosphates (dNTP)
Documentation of test results, 468–469
  gene nomenclature, 469
  gene sequencing results, 469
  reporting results, 469–470
  resequencing, 469
  summary sheet or database, 470
Domains, 38
Dominant, 438
Dominant-negative phenotype, 347, 348*f*
Dot blot, 133–134, 134*f*, 209
Double helix, 4*b*, 5*f*, 7*b*, 9*b*
Double-strand break, 16, 19
Double-stranded DNA, 8, 102*f*
Dual-fusion probes, 191
Dye blobs, 232
Dye comigration, 109*t*
Dye primer, 230, 231*f*
Dye terminator, 230, 231*f*

E

E site, 49
EBNA. *See* EBV nuclear antigen (EBNA)
EBV. *See* Epstein-Barr virus (EBV)
EBV nuclear antigen (EBNA), 320
EGF. *See* Epidermal growth factor (EGF)
EGFR. *See* Epidermal growth factor receptor (EGFR)
EIA. *See* Enzyme immunoassays (EIA)
Electrodes, 99
Electroendosmosis, 99
Electropherogram, 232, 233*f*, 276*f*, 278*f*
Electrophoresis, 231–232
  agarose, 98
  anode, 98
  defined, 98
  electropherogram, 232
  measurement, 90–91
  polyacrylamide, 98
Electrophoresis capillary replacement, 459*f*
Electrophoresis equipment, 106–109
  comb, 106
  shark tooth combs, 107
  submarine gels, 106
  wells, 106
Electrophoretic transfer, 120, 121*f*
ELISA. *See* Immunosorbent assays (ELISA)
Elongation, transcription, 28–29, 28*f*
Emulsion PCR, 166–167
End labeling, 126
Endonucleases, 14
Endpoint analysis, 159
Engraftment, 284
Enhancers, 63
Enterobacterial repetitive intergenic consensus (ERIC), 333, 335*f*
Enzymatic and chemical cleavage methods, 215–217
Enzyme adaptation (induction), 61, 61*b*
Enzyme immunoassays (EIA), 201–202, 202*b*
Enzyme induction, 63
Enzyme repression, 63
Enzymes metabolizing DNA, 14–19
  DNA metabolizing enzymes, 17–19
  helicases, 18–19
  ligase, 17
  methyltransferases, 19
  nucleases, 17–18
  restriction enzymes, 14–17, 15*t*
Epidemic, 329
Epidemiological typing methods, 330*t*
Epidermal growth factor (EGF), 374*f*
Epidermal growth factor receptor, *EGFR* (7p12), 373–375
Epigenetic factor classification, 69
Epigenetic profiles, 294
Epigenetic regulation, 66
Epigenetics, defined, 65
Episomal nature of the resistance factor, 25–26

Episomes, 26
Epitopes, 123, 202
Epstein-Barr virus (EBV), 320
ER stress, 51*b*
ERIC. *See* Enterobacterial repetitive intergenic consensus (ERIC)
ESI spectrophotometry, 205–206, 206*f*
Ethidium bromide, 90, 91*f*
Euchromatin, 67*b*, 183
Eukaryote(s), 29, 30*f*, 58*t*, 63*f*
Euploid organisms, 181
E-value, 248
Ewing sarcoma, *EWS* (22q12), 376–377, 377*t*, 378*f*
Exclusion, 265, 272, 276
Exons, 32
Exonuclease II, 18
Exonucleases, 12, 17
Expect value, 248
Explosive materials, 450*f*, 465*f*
Expression arrays, 136
Extended haplotype, 281*b*
Extended MHC locus (xMHC), 418
Extension/termination assay, 146*b*
Extraction with chelating resin, 86

F

Factor V Leiden mutation, 349, 352, 352*f*
Factura, 235*t*
False negative, 153
False positive, 306
FASTA, 235*t*
Fertility factor, 21, 22*f*
F factor, 23
Fidelity, 14, 150
Field inversion gel electrophoresis (FIGE), 100, 100*f*
FIGE. *See* Field inversion gel electrophoresis (FIGE)
FISH analysis, 192*f*
Flammable materials, 465*f*
Flemming, Walther, 3
Fluorescence, TaqMan signal, 162*f*
Fluorescence in situ hybridization, 190–195
  interphase FISH, 190–195, 191*f*
  metaphase FISH, 193–194
Fluorescent AFLP analysis, 335*f*
Fluorescent dyes, 109–110
Fluorescent resonance energy transfer, 162–163, 163*f*, 210–211, 210*f*, 211*f*
Fluorescent sequencing chemistries, 231*f*
Fluorometry, 93–94
Follicular lymphoma, 406*t*
FR. *See* Framework region (FR)
Fragile X chromosome, 359*f*
Fragile X syndrome (FXS), 358–360, 359*f*, 360*f*
Fragment analysis, 286
Frameshift mutation, 200, 201*b*, 219
Framework region (FR), 393
FRET. *See* Fluorescent resonance energy transfer), 162–163, 163*f*, 210–211, 210*f*, 211*f*
FRET probes, 163*f*
Full chimerism, 285
Fume hoods and laminar flow hoods, 462–463
Fungi
  detection of, 323–324
  DNA isolation, 80
Fungicidal, 324
Fungistatic antimicrobial agents, 324
Fusion gene, 401
FXS. *See* Fragile X syndrome (FXS)

G

G1 checkpoint, 371
G2 checkpoint, 371
Gain-of-function mutations, 347, 371
Gamete, 14
Gamow, George, 44, 44*b*
Gap, 249*t*, 252
Gap-filling DNA polymerase, 11*b*
Gas chromatography, 204–205, 205*f*
G-banding, 185, 185*b*, 185*f*, 186*f*
Gel, 99
Gel electrophoresis, 155*f*, 461*f*
Gel loading, 108–109
Gel mobility shift assay, 139, 139*f*
Gel photography, cameras, 462*f*
Gel systems, 99–102
  agarose gels, 99–101
  buffers, 99
  capillary electrophoresis, 102–104
  field-inversion gel electrophoresis, 100
  polyacrylamide gels, 101–102
  pulsed-field gel electrophoresis, 100–101
GenBank, 248, 249*t*
Gender identification, 270, 270*f*
Gene, 43, 43*f*
Gene expression, 27, 58
Gene mutations, 181, 199–222
  allele-specific primer amplification, 214*f*
  allelic discrimination, 215*f*
  antigen detection, 202*f*
  bead array technology, 213*f*
  cleavase assay, 216, 218*f*
  cleavase single-color assay, 218*f*
  curve analysis, homozygous mutant, 209*f*
  detection of gene mutations, 201–218
    biochemical methods, 201–207
    chemical cleavage, 215–216
    enzymatic and chemical cleavage methods, 215–217
    heteroduplex analysis with single-strand specific nucleases, 216
    methods, 217–218
    nonisotopic RNase cleavage assay, 216, 217*f*
    nucleic acid analyses, 207–218
    restriction fragment length polymorphisms, 215–216, 215*f*, 216*f*
    sequencing-based methods, 213–215
  ESI spectrophotometry, 205–206, 206*f*
  frameshift mutation, 200, 201*b*, 219
  gas chromatography, 205*f*
  gene names, 219–220
  gene variant nomenclature, 218–219
  heteroduplex analysis, 212*f*

- Gene mutations *(cont'd)*
  - liquid chromatography, 205*f*
  - MALDI spectrophotometry, 205–206, 206*f*
  - MALDI-TOF spectrophotometry, 206, 206*f*
  - melt-curve analysis, 209*f*
  - multiplex allele-specific PCR, 214*f*
  - multiplex PCR, 216, 216*f*
  - mutation detection methodologies, 208*t*
  - NIRCA analysis, 217*f*
  - PCR-RFLP, 215–216, 215*f*
  - point mutations, 200, 200*t*
  - sequence-specific primer amplification, 213*f*
  - single-strand conformation polymorphism, 208*f*
  - types of, 200–201
- Gene names, 219–220
- Gene nomenclature, 469
- Gene panels, 240
- Gene rearrangements, 387–397
  - banding patterns, 396–397
  - defined, 387
  - detection of clonality, 391
  - immunoglobulin heavy-chain gene rearrangement in B cells, 387–389, 388*f*
  - immunoglobulin light-chain, 394, 395*f*
  - immunoglobulin light-chain gene rearrangement in B cells, 389–390, 389*f*
  - molecular analysis of immunoglobulin heavy-chain gene clonality, 391–394, 392*f*, 393*f*, 394*f*
  - T-cell receptor, 390–391, 390*f*, 390*t*, 391*f*, 394–396, 395*f*, 396*f*
  - V(D)J recombination, 387, 387*f*
- Gene sequencing results, 469
- Gene variant nomenclature, 218–219
- Gene/chromosomal mutations, solid tumors, 372–387
- Genes conferring resistance to antimicrobial agents, 326*t*
- Genetic code, 37–46
  - alpha helix, protein, 42, 42*f*
  - amino acid(s), 38–43
    - biosynthetic groups of, 40*t*
    - structures, 39*f*
  - amino acyl tRNA synthetases, 47*f*
  - amino terminal, 38
  - aminoacyl tRNA synthetases, 46
  - anticodon, 45–47
  - beta-pleated sheets, 42
  - carboxy terminal, 38
  - chaperones, 49
  - chromosome, 43
  - codons, 45, 45*f*
  - complement, 44*b*
  - conjugated proteins, 43
  - consensus, 42*b*
  - domains, 38
  - E site, 49
  - ER stress, 51*b*
  - extracellular domain, 38
  - gene, 43
  - glycoproteins, 43
  - isodecoders, 46
  - leucine zipper, 42*b*
  - lipoproteins, 43
  - metalloproteins, 43
  - monomer, 43
  - nonpolar, 38
  - nonsense codons, 45
  - nonsense-mediated decay, 51
  - oligomers, 43
  - P site, 47, 49
  - peptide, 38
  - peptide bond, 38, 41*f*
  - peptidyl transferase, 49
  - polar, 38
  - polypeptides, 38
  - primary structure, 41
  - prions, 43
  - prosthetic group, 43
  - protein synthesis, 50*f*
  - proteins, 38, 43–44, 45*f*
  - proteome, 38
  - proximal regulatory sequences, 43
  - pyrrolysine, 38*b*
  - quaternary structure, 43
  - random coils, 42
  - regulatory sequences, 43
  - release factors, 51
  - ribosomal binding site, 47
  - ribosome(s), 47, 48*f*
  - RNA Tie Club, 44*b*
  - secondary structure, 42
  - selenocysteine, 38*b*
  - selenoproteins, 38*b*
  - side chain, 38
  - structural sequences, 43
  - suppressor tRNAs, 47*b*
  - tertiary structure, 42–43
  - translation, 46–51
    - amino acid charging, 46–47
    - protein synthesis, 47–51, 50*f*
  - transmembrane proteins, 40*f*
  - tRNA charging, 46
  - zinc finger, 42*b*
  - zwitterions, 38
- Genetic concordance, 272
- Genetic mapping with RFLPs, 264
- Genetic profiling, 265–266, 266*b*
- Genital tract organisms, 312*t*
- Genome, chromosomal mutations, 186–195
  - fluorescence in situ hybridization, 190–195
    - interphase FISH, 190–195, 191*f*
    - metaphase FISH, 193–194
  - karyotyping, 186–190
    - balanced translocation, 187
    - chromosome spread, 186
    - defined, 186
    - deletion, 187
    - derivative chromosome, 188
    - descriptive abbreviations, 191*t*
    - insertion, 187–188
    - inversions, 188
    - isochromosome, 188, 189*f*
    - microdeletions, 187

        - mitogen, 186
        - paracentric inversions, 188
        - pericentric inversions, 188
        - reciprocal translocations, 186
        - ring chromosome, 188
        - robertsonian, 187
        - translocations, 186
        - unbalanced translocation, 187
- Genome mutations, 181, 346*t*
- Genome sequencing, 12*b*
- Genomes, human viruses, 314*t*
- Genomic amplification methods, 165–168
    - arbitrarily primed PCR, 167–168, 169*f*
    - bridge PCR, 167, 168*f*
    - comparative genomic arrays, 166
    - emulsion PCR, 166–167
    - multiple displacement amplification, 166
    - randomly amplified polymorphic DNA, 167
    - solid-phase emulsion PCR, 166–167, 167*f*
    - surface amplification, 167, 168*f*
    - whole-genome amplification, 165, 165*f*, 166, 166*f*
- Genomic array technology, 134–138
    - macroarrays in, 134
    - microarrays in, 134–138, 135*f*
    - microelectronic arrays, 135–138
- Genomic DNA fragments, 118*f*
- Genomic imprinting, 345, 362–363
- Genomics, 136*b*
- Genotype, 180
- Germline, 387–388
- Gill, Peter, 275*b*
- Gloves, 450*f*
- Glycopeptide antibiotic resistance, 328–329
- Glycoproteins, 43
- Gonadal mosaicism of new mutations, 348*b*, 355–356, 356*f*
- Gorer, P., 418*b*
- G-proteins, 376
- Graft failure, 285
- Graft-versus-host disease (GVHD), 284, 418
- Graft-versus-tumor (GVT), 284
- GRAIL, 235*t*
- GTPase-activating proteins, 376
- Guanine, 3
- Guanosine, 4
- GVHD. *See* Graft-versus-host disease (GVHD)
- GVT. *See* Graft-versus-tumor (GVT)
- Gyrases, 19

H

- HAART. *See* Highly active antiretroviral therapy (HAART)
- Hairpin, 29, 59
- Hairy cell leukemia, 406*t*
- Haploid genes, 180
- Haplotype, 278, 420, 421*f*
- Haplotype diversity, 279
- HapMap project, 146*b*, 253, 290–291, 290*f*, 291*b*
- Haptens, 125
- Hardy-Weinberg equilibrium, 274*b*
- Harvey rat sarcoma viral oncogene homolog, *H-ras* (11p15), 375–376, 375*f*, 375*t*
- Hazard labels, 466*f*
- HCV. *See* Hepatitis C virus (HCV)
- HeLa cervical carcinoma, 33*t*
- Helicases, 11*b*, 18–19, 37
- Hemachromatosis, 352–353, 353*f*, 354*f*
- Hematological malignancies
    - gene mutations, 404–405
        - CCAAT/enhancer-binding protein, alpha (*CEBPA*), 404
        - FMS-related tyrosine kinase 3, 404–405
        - Janus kinase 2, 405
        - nucleophosmin/nucleoplasmin family, member 1 (*NPM1*), 404
    - molecular analysis, 387–406
    - molecular targets
        - kappa deleting element, 389
        - T-cell receptor gene rearrangement, 390–391
        - trimming, 391
    - mutations
        - lymphoid, 397–400
        - myeloid, 401–405
- Hematopoietic stem cells, 283
- Hemimethylated, 19
- Hemizygous for X-linked genes, 348
- Hepatitis C virus, 321–322
- Herpes simplex virus (HSV), 320
- Herpes virus (VZV), 320–322
    - cytomegalovirus, 320–321
    - Epstein-Barr virus, 320
    - herpes simplex virus, 320
- Hershey, Alfred, 23
- Heterochromatin, 67*b*, 183
- Heteroduplex analysis, 211, 212*f*, 216
- Heteroduplexes, 98
- Heterologous extrinsic controls, 305
- Heterologous intrinsic controls, 305
- Heteronuclear RNA, 30*f*, 32
- Heterophagy, 349
- Heteroplasmy, 293
- Heterosis, 19
- Heterozygosity loss, 385, 386*f*
- Heterozygous, 263
- HFE protein, 353*f*
- Hfr, 23
- Hierarchal shotgun sequencing, 251, 252*f*
- High-density oligonucleotide arrays, 136, 212
- Highly active antiretroviral therapy (HAART), 317
- High-performance liquid chromatography (HPLC), 204
- High-positive controls, 457
- High-resolution banding, 186
- Hirschsprung disease, 381
- Histocompatibility antigens, 437
- Histone(s), 181–183
    - code of, 67*b*, 67*t*, 68*f*
    - modification of, 65–66, 66*f*, 68*f*
- HIV. *See* Human immunodeficiency virus (HIV)
- HLA. *See* Human leukocyte antigens (HLA)
- Hoagland, amino acid charging studies, 46
- Hodgkin disease, 370
- Hoechst 33258, 94
- Holding and storage requirements, 451
- Holley, Robert, 35*b*

Holocentric chromosomes, 184*b*
Holoenzyme, 11–12, 35*f*
Homoduplexes, 211
Homogeneous MassExtend, 146*b*
Homologous, 66
Homologous extrinsic controls, 305
Homologous partners, 207–208
Homologous sequences, 9*f*
Homology, 248, 249*t*
Homozygous, 263
Horizontal gel electrophoresis, 98*f*
Horizontal submarine gel, 106*f*
Hot-start PCR, 154–155
HPLC. *See* High-performance liquid chromatography (HPLC)
HPV. *See* Human papillomavirus (HPV)
HSV. *See* Herpes simplex virus (HSV)
HUGO. *See* Human Genome Organization (HUGO)
Human epidermal growth factor receptor 2, *HER2/neu/erb-b2 1* (17q21.1), 372–373
Human genome, 180
Human Genome Organization (HUGO), 219, 270*b*
Human Genome Project, 250–254, 251*t*
Human Haplotype Mapping Project, 146*b*, 253, 290–291, 290*f*, 291*b*
Human identification, 260–300
  allele peak locations, 269*f*
  amelogenin locus, 270*f*
  autologous bone marrow transplant, 283*f*
  bone marrow engraftment testing, DNA polymorphisms, 283–289
    allogeneic transplant, 283
    autologous transplant, 283
    chimera, 284
    fragment analysis, 286
    full chimerism, 285
    graft-versus-host disease, 284
    graft-versus-tumor, 284
    hematopoietic stem cells, 283
    informative locus, 285
    linkage analysis, 282–283, 282*f*
    matched unrelated donor, 284
    mixed chimerism, 285
    mosaic, 284*b*
    myeloablative, 284, 287
    noninformative loci, 285
    post-transplant engraftment testing, 287–289, 287*f*
    pretransplant STR testing, 285–287
    split chimerism, 286*b*
    stutter, 286–287
  capillary gel electrophoresis, 269*f*
  electropherogram, 276*f*, 278*f*
  epigenetic profiles, 294
  linkage analysis, 282–283, 282*f*
  microvariant allele, 272*f*
  mitochondrial DNA polymorphisms, 291–293, 291*f*
  PAGE analysis, 285*f*
  paternity tests, 276*t*
  polymorphisms, 262*t*
  protein-based identification, 293–294
  quality assurance for surgical sections using STR, 289, 289*f*
  RFLP in, 262–266, 264*f*
  short tandem repeat TH01, 267*f*, 268*f*
  single nucleotide polymorphisms, 289–291, 290*t*
  Southern blot, 262*t*
  STR locus information, 271*t*
  STR typing by PCR, 266–278
    allelic ladders, 268
    discrete allele systems, 268
    gender identification, 270, 270*f*
    internal size standards, 268
    microsatellites, 267
    microvariants, 267
    mini-STR, 268
    short tandem repeats, 267
    STR analysis, 268–278
    STR nomenclature, 270
    test results analysis, 270, 272, 272*b*, 272*t*, 273*b*
  tandem repeat, 265*f*
  types of polymorphisms, 261
  Y-STR, 278–282
    autosomal STRs, 278
    genotypes, 281*t*
    locus information, 280*t*–281*t*
    matching with Y-STRs, 279, 281–282
    paternal lineage test, 279
Human immunodeficiency virus (HIV), 314, 317–319
  genotyping, 319
  HIV genotyping, 319
  limit of detection, 317–318
  primary resistance mutations, 319
  secondary resistance mutations, 319
  viral load, 317, 318*t*
Human leukocyte antigens (HLA)
  alleles identified serologically and by DNA sequence, 426*t*
  classes, 418, 419*f*
  defined, 418
  HLA and KIR gene interactions, 438*t*
  nomenclature, 420, 422–425
    ambiguity, 425
    null allele, 425
    resolution of allele detail, 425
    split specificities, 422
    synonymous changes, 425
  polymorphisms, 420–425
    allele, 420, 421–422
    haplotype, 420, 421*f*
    type, 420
  serologically defined specifications, 422*t*–424*t*
  test discrepancies, 436–437
  type, 420
  typing, 427–428
Human papillomavirus (HPV), 322
Humoral sensitization, 428
Hungerford, David, 188*b*
Huntingtin repeat expansion, 361*f*
Huntington disease, 361
Hybrid, 157
Hybrid capture assays, 173, 173*f*
Hybrid resistance, 437
Hybrid vigor, 14
Hybridization, 9

- Hybridization conditions, stringency, 128–129
  - C_0 t value, 128
  - C_0 t½, 128
  - melting temperature, 128, 128*f*
  - sequence complexity, 128
- Hybridization technologies, 116–121, 117*t*
  - northern blot, 122
  - Southern blot, 117–121
  - western blot, 122–123
- Hybridization-based methods
  - allele-specific oligomer hybridization, 209
  - array technology, 212–213
  - heteroduplex analysis, 211
  - melt-curve analysis, 209–211
  - single-strand conformation polymorphism, 207–209
- Hybridomas, 126
- Hydrogen bonds, 7
- Hypervariable region I, 291, 292*f*
- Hypervariable region II, 291, 292*f*

I

- Iatrogenic, 330
- Identification, microorganisms, 301–343
  - AFLP analysis, 334*f*
  - amplification controls, 306*f*
  - antimicrobial action sites, 325*f*
  - antimicrobial agents, 324–329
    - bacteriocidal, 324
    - bacteriostatic, 324
    - fungicidal, 324
    - fungistatic, 324
    - mode of action, 325*t*
    - molecular detection of resistance, 327–329
    - resistance to, 302, 325–327
  - bacteria, 309–313
    - respiratory tract pathogens, 309–311
    - urogenital tract pathogens, 311–313
  - beta-lactamase, reaction, 328*f*
  - coagulase-negative *Staphylococcus*, 331*f*
  - epidemiological typing methods, 330*t*
  - fluorescent AFLP analysis, 335*f*
  - fungi, 323–324
  - genital tract organisms, 312*t*
  - genomes, human viruses, 314*t*
  - HIV viral loads, nucleic acid amplification methods, 318*t*
  - microorganisms, genes conferring resistance to antimicrobial agents, 326*t*
  - molecular detection, microorganism, 308–324
  - molecular epidemiology, 329–337
    - epidemic, 329
    - molecular strain typing, 330–336
    - pandemic, 329
    - typing method comparison, 336–337, 337*t*
  - nucleic acid amplification tests, viruses, 315*t*–317*t*
  - nucleic acid test, target sequences, 307*f*
  - parasites, 324
  - penicillin structure, 328*f*
  - PFGE pattern interpretation, 332*t*
  - quality control, 305–306
  - RAPD gel results, 333*f*
  - resistance mechanisms, 326*t*
  - respiratory tract organisms, 310*t*
  - reverse transcriptase, 302
  - ribosomal RNA genes, 336*f*
  - sample preparation, 304–305
  - sequence targets, 307
  - specimen collection, 302–304
  - specimen transport systems, 303*t*
  - Swab Extraction Tube System, 303*f*
  - vancomycin structure, 329*f*
  - vancomycin-resistant *S. aureus*, 327*f*
  - viruses, 313–323
    - BK/JC viruses, 323
    - cytopathic effect, 313–314
    - hepatitis C virus, 321–322
    - herpes viruses, 320–321
    - human immunodeficiency virus, 314, 317–319
    - human papillomavirus, 322
    - load measurement, test performance, 318*t*
    - mass spectrometry, 323
    - mycology, 323–324
    - nucleic acid amplification tests, 315*t*–317*t*
    - respiratory viruses, 322–323
- Identity, 249*t*
- Idiopathic congenital central hypoventilation syndrome, 361–362
- Immobilized, 85
- Immunoassays, 201–202
- Immunoglobulin, 389*f*
- Immunoglobulin heavy chain
  - rearrangement of, 388*f*, 391–394, 392*f*, 393*f*
  - rearrangement of B cells, 387–389, 388*f*
- Immunoglobulin light chain gene rearrangement
  - B cells, 389–390, 389*f*
  - clonality detection, 394
  - kappa locus, 395*f*
  - lambda locus, 395*f*
- Immunohistochemistry, 201, 202–204, 203*f*
- Immunosorbent assays (ELISA), 202*b*
- Imprinting, 66
- Imprinting, genomic, 362–363
- In silico, 248
- In vitro analytical test (IVAT), 465
- In vitro diagnostic (IVD), 464–465
- Inclusion, 265
- IncRNA, 71–72, 73*f*
- Indexing, 240, 241*f*
- Indirect nonradioactive detection, 130*f*
- Inducible transcription, 30
- Induction (enzyme adaptation), 61, 61*b*
- Informative locus, 285
- Informative STR alleles, 286*f*
- Inheritance of alleles, 282, 283*f*
- Inherited disease detection, 344–368
  - autosomal-dominant transmission, 347*f*
  - autosomal-recessive mutations, 348*f*
  - chromosomal abnormalities, 345–346
  - chromosomal mutations, 253*t*
  - cystic fibrosis, 353–354, 354*f*
  - cystic fibrosis transmembrane conductance, 353, 354*f*
  - cytochrome P-450 enzymes, 354–355, 355*f*

Inherited disease detection *(cont'd)*
  factor V Leiden mutation, 349, 352, 352*f*
  fragile X chromosome, 359*f*
  fragile X syndrome, 359*f*
  genome mutations, 346*t*
  genomic imprinting, 362–363
  gonadal mosaicism, 356*f*
  hemachromatosis, 352–353, 353*f*, 354*f*
  HFE protein, 353*f*
  Huntingtin repeat expansion, 361*f*
  limitations to molecular testing, 363
  lysosome, 351*f*
  methylenetetrahydrofolate reductase enzyme, 352, 353*f*
  mitochondrial deletion, 358*f*
  mitochondrial disorders, 357*t*
  mitochondrial genome, 356*f*
  mitochondrial mutations, 355*f*, 356–357
  molecular basis, inherited diseases, 345, 345*b*
  multimeric proteins, 347, 348*f*
  NARP mitochondrial point mutation, 358*f*
  nonpolyglutamine expansion disorders, 363*t*
  pedigree, 347*f*
  prothrombin, 349–350, 351*t*, 352
  sequence-specific PCR, 352*f*
  single-gene disorders, 355–363
    inheritance patterns, 346–349
    molecular basis, 349–355
    molecular methods, 350*t*
    nonclassical patterns, 355–363
  storage diseases, 351*t*
  thrombosis, 351*t*
  triplet-repeat expansion, 359*f*, 361, 362*f*
  uniparental disomy, 362–363
  X-linked recessive diseases, 348*f*
Initiation, transcription, 28
Ink-jet technology, 135*f*
Inorganic DNA isolation, 83, 83*f*
Insertion, 187–188
Instrument maintenance, 459–463
  block thermal cyclers, 460
  calibration verification, 457, 463
  calibrations, 456–457, 463
  cameras and detectors, 462
  centrifuges, 461
  coefficient of variance, 457, 461
  fume hoods and laminar flow hoods, 462–463
  microcentrifuge speed controls, 461
  power supplies, 461
  refrigerators and freezers, 459–460
  reportable range, 455, 463
Intercalating agents, 8*b*, 109–110
Interface, 239, 249*t*
Internal controls, 305, 457
Internal labeling, 226
Internal size standards, 268
Internal transcribed spacer (ITS), 310, 334
Interphase FISH, 190–195, 191*f*
Interpretation of results, 132–133
Interspersed repetitive elements, 333–334
Intervening sequences, 31, 31*b*
Introns, 32
Invader assay, 174*f*
Inversions, 188
Investigational use only (IUO), 465
Isochromosome, 188, 189*f*
Isocratic, 204
Isodecoders, 46
Isolation of DNA, 79–87
  bacteria and fungi, 80
  inorganic isolation methods, 83
  isolation of mitochondrial DNA, 86
  lipoproteins, 81
  mitochondrial DNA, 86
  nucleated cells in suspension, 80
  organic isolation methods, 81–83
  plasma, 80–81
  preparing sample, 79–81
  proteolytic lysis, 85–86
  salting out, 83
  solid-phase isolation, 84–85, 84*f*
  tissue samples, 81
  viruses, 80
Isolation of RNA, 87–90
  diethyl pyrocarbonate, 87*b*
  extraction of total RNA, 87
  intracellular, 88
  microfuge, 88–89
  polyA RNA, 89–90
  proteolytic lysis of fixed material, 89
  RNAse-free, 87
  specimen collection, 87
  specimen sources, 90*t*
  total RNA, 87
Isotype, 388*b*
ITS. *See* Internal transcribed spacer (ITS)
IUB universal nomenclature, 248–250, 250*t*, 251, 252*f*, 253*t*
IUO. *See* Investigational use only (IUO)
IVAT. *See* In vitro analytical test (IVAT)
IVD. *See* In vitro diagnostic (IVD)

J

Jacob, Francois, 23
Janus kinase 2, 405
Jeffrey, Alec, 275*b*
Joining reaction, 17*b*–18*b*

K

Kammerer, Paul, 72*b*–73*b*
Kappa deleting element (KDE), 389
Karyotyping, 183, 186–190
  balanced chromosome, 187
  chromosome spread, 186
  defined, 186
  deletion, 187
  derivative chromosome, 188
  descriptive abbreviations, 191*t*
  insertion, 187–188
  inversions, 188
  isochromosome, 188, 189*f*
  microdeletions, 187

  - mitogen, 186
  - paracentric inversions, 188
  - pericentric inversions, 188
  - reciprocal translocations, 186
  - ring chromosome, 188
  - robertsonian, 187
  - showing balanced reciprocal translocation, 189*f*
  - translocations, 186
  - unbalanced translocation, 187
- KDE. *See* Kappa deleting element (KDE)
- Khorana, H. Gobind, 17*b*, 43, 45
- Killer cell immunoglobulin-like receptors, 437–438
- Kinetochore, 183, 184*f*
- King, Mary Claire, 264*b*
- Kinship index, 277
- KIR and HLA gene interactions, 438*t*
- KIR gene cluster, 438*f*
- Kirsten rat sarcoma viral oncogene homolog, K-ras (12p 12), 375–376, 375*f*, 375*t*, 376*f*
- Klenow fragment, 12
- Kornberg, Arthur, 12*b*
- Kornberg, Sylvy, 12*b*

L

- Labeling, probes, 126
- Labeling of reagents, 466–468
- Laboratory techniques, 259–472. *See also* Molecular laboratory quality control
  - DNA polymorphisms and human identification, 260–300
  - DNA-based tissue typing, 417–445
  - inherited disease molecular detection, 344–368
  - microorganism detection and identification, 301–343
  - molecular oncology, 369–416
  - quality assurance and quality control, 446–472
  - testing summary, 439–440
- Laboratory-developed tests, 451
- Lac operon, 62*f*
- Lagging strand, 10
- Laminar flow hoods, 462*f*
- Leading strand, 10
- Leder, Philip, 43, 44
- Lederberg, Joshua, 23*b*
- *Legionella pneumophila*, 309
- Lehman, I. Robert, 12*b*
- Leucine zipper, 42*b*
- Leukemia
  - characteristic, 370
  - chromosomal abnormalities, 406*t*
  - gene rearrangements, 387–397
  - molecular analysis, 388–391
- Leukocyte receptor cluster, 438
- Library, 240
- Ligase, 13, 17
- Ligase chain reaction, 168–169, 169*f*
- Likelihood ratio, 274
- Limit of detection, 305, 317–318
- Limitations to molecular testing, 363
- Linearity, 454*t*
- LINEs. *See* Long interspersed nucleotide sequences (LINES)
- Linkage analysis, 282–283, 282*f*
- Linkage disequilibrium, 282
- Linkage equilibrium, 274
- Linker DNA, 181, 183*f*
- Lipoproteins, 43, 81
- Liquid biopsy, 81, 386–387
- Liquid chromatography, 205*f*
- Local alignment, 249*t*
- Locked nucleic acids, 123
- Locus, 263
- Locus genotype, 272
- Locus-specific brackets, 273*b*
- Locus-specific RFLP, 332
- Long interspersed nucleotide sequences (LINEs), 261
- Long noncoding, 71–72
- Loss of heterozygosity (LOH), 385, 386*f*
- Loss-of-function, 347, 371–372
- Low-copy-number analysis, 275*b*
- Low-positive controls, 457
- Lymphocytic leukemia, 406*t*
- Lymphoid malignancy mutations, 397–400
- Lymphoma
  - ALK proto-oncogene 2p23.1, 382
  - chromosomal abnormalities, 406*t*
  - defined, 370
  - gene rearrangements, 387–397
  - molecular analysis, 388–391
  - non-Hodgkin, 370
- Lynch syndrome, 382
- Lyon hypothesis, 345*b*
- Lysosomal storage diseases, 349
- Lysosome, 349, 351*f*

M

- M13 universal primer, 226*b*
- Macroarrays, 134
- Major breakpoint region, 397
- Major groove, 8*b*
- Major histocompatibility complex (MHC)
  - disease association, 438–439
  - genes, 419*t*
  - MHC locus, 418–420
    - extended MHC locus, 418
    - mixed lymphocyte culture, 418
  - nonconventional antigens, 437
- MALDI spectrophotometry, 205–206, 206*f*
- MALDI-TOF spectrophotometry, 206, 206*f*
- Male karyotype, 187*f*
- Mantle cell lymphoma, 406*t*
- Manual dideoxy sequencing, 225*f*
- Manual sequencing, 224–229
  - chemical sequencing, 224–225
  - dideoxy chain termination (Sanger) sequencing, 225–229
    - alkaline phosphatase, 227*b*
    - dideoxynucleotide, 226, 227*f*
    - direct sequencing, 225*f*, 226*f*
    - internal labeling, 226
    - M13 universal primer, 226*b*
    - sequencing ladder, 228, 229*f*, 230
  - Maxam-Gilbert sequencing, 224–225
- Marker chromosomes, 346

Mass spectrometry, 205–207, 323, 336
MassExtend, 146*b*
Master mix, 152
Match probability, 273
Matched unrelated donor (MUD), 284
Matching of profiles, 273–275
Matching with Y-STRs, 281–282
  discriminatory capacity, 279
  haplotype diversity, 279
  surname test, 282
Matchmaker, 235*t*
Maxam-Gilbert sequencing, 224–225, 224*t*, 225*f*. *See also* Chemical sequencing
Matrix Assisted Laser Desorption/Ionization – Time of Flight (MALDI-TOF) spectrophotometry, 206, 206*f*
MCA. *See* Melt-curve analysis (MCA)
Measurement
  analyte measurement range, 454*t*
  electrophoresis, 90–91
  fluorometry, 93–94
  microfluidics, 94
  nucleic acid quality and quantity, 90–94
  spectrophotometry, 92–93
  viral load measurement, test performance, 318*t*
mecA (methicillin resistance), 174
Meiosis, 20
Melt-curve analysis (MCA), 209–211
  fluorescent resonance energy transfer, 210–211, 210*f*, 211*f*
  PCR products, 209*f*
Melting temperature, 128, 128*f*
Membrane types, 119–120
  anode, 120
  cathode, 120
  electrophoretic transfer, 120, 120*f*
  prehybridization, 121–122
  vacuum transfer, 121
MEN. *See* Multiple endocrine neoplasia (MEN) syndromes, 381
Messenger RNA (mRNA), 29–33
  function, 27, 58
  processing, 31
  splicing, 31–33, 32*f*
Metabolizing enzymes, RNA, 36–37
Metacentric chromosome, 184, 184*f*
Metalloproteins, 43
Metaphase chromosomes, 182*f*, 194*f*
Metaphase FISH, 193–194
  spectral karyotyping, 193
  whole chromosome paints, 193, 194*f*
Metastasis, 370
Methicillin and oxacillin, 328, 328*f*
Methicillin resistance (*mecA*), 174
Methicillin-resistant *S. Aureus*, 326
Methods of gene mutation detection, 217–218
Methylenetetrahydrofolate reductase, 352, 353*f*
Methyltransferases, 19
mHag. *See* Minor histocompatibility antigens (mHag)
MHC. *See* Major histocompatibility complex (MHC)
MIC. *See* Minimum inhibitory concentration (MIC)
Microarray(s), 134–138, 135*f*
Microcentrifuge speed controls, 461
Micrococcal nuclease, 18
Microdeletions, 187
Microelectronic arrays, 134
Microfluidics, 94
Microorganism detection, 301–343
  AFLP analysis, 334*f*
  amplification controls, 306*f*
  antimicrobial action sites, 325*f*
  antimicrobial agents, 302, 324–329
    bacteriocidal, 324
    bacteriostatic, 324
    fungicidal, 324
    fungistatic, 324
    mode of action, 325*t*
    molecular detection of resistance, 327–329
    resistance to, 302, 325–327
  bacteria, 309–313
    respiratory tract pathogens, 309–311
    urogenital tract pathogens, 311–313
  beta-lactamase, reaction, 328*f*
  coagulase-negative *Staphylococcus*, 331*f*
  epidemiological typing methods, 330*t*
  fluorescent AFLP analysis, 335*f*
  fungi, 323–324
  genital tract organisms, 312*t*
  genomes, human viruses, 314*t*
  HIV viral loads, nucleic acid amplification methods, 318*t*
  microorganisms, genes conferring resistance to antimicrobial agents, 326*t*
  molecular detection, microorganism, 308–324
  molecular epidemiology, 329–337
    epidemic, 329
    molecular strain typing, 330–336
    pandemic, 329
    typing method comparison, 336–337, 337*t*
  nucleic acid amplification tests, viruses, 315*t*–317*t*
  nucleic acid test, target sequences, 307*f*
  parasites, 324
  penicillin, 328*f*
  penicillin structure, 328*f*
  PFGE pattern interpretation, 332*t*
  quality control, 305–306
  RAPD gel results, 333*f*
  resistance mechanisms, 326*t*
  respiratory tract organisms, 310*t*
  reverse transcriptase, 302
  ribosomal RNA genes, 336*f*
  sample preparation, 304–305
  sequence targets, 307
  specimen collection, 302–304
  specimen transport systems, 303*t*
  Swab Extraction Tube System, 303*f*
  vancomycin structure, 329*f*
  vancomycin-resistant *S. aureus*, 327*f*
  viral load measurement, test performance, 318*t*
  viruses, 313–323
    BK/JC viruses, 323
    cytopathic effect, 313–314
    DNA isolation, 80
    DNA template, 149

        hepatitis C virus, 321–322
        herpes viruses, 320–321
        human immunodeficiency virus, 314, 317–319
        human papillomavirus, 322
        mass spectrometry, 323
        mycology, 323–324
        nucleic acid amplification tests, 315*t*–317*t*
        respiratory viruses, 322–323
MicroRNA (mRNA), 70
Microsatellites
    instability of, 382, 384–385, 385*f*
    STR typing by PCR, 267
Microtome, 202
Microvariant allele, 272*f*
Microvariants, 267
Miescher, Johann Friedrich, 3, 3*b*–4*b*
Migration of DNA, 25
Minimal haplotype, 281*b*
Minimum inhibitory concentration (MIC), 327
Minisatellites, 267
Mini-STR, 268
Minor cluster region, 397
Minor groove, 8*b*
Minor groove-binding dyes, 110
Minor histocompatibility antigens (mHag), 437
miRNA. *See* MicroRNA (mRNA)
miRNA silencing, 72*f*
Mismatch repair (MMR), 382, 383*t*, 384*f*
Mismatches, 7
Mispriming, 148, 148*f*
Mitochondrial deletion, 356, 358*f*
Mitochondrial disorders, 357*t*
Mitochondrial DNA polymorphisms, 291–293
    Anderson reference, 291
    Cambridge reference sequence, 291
    hypervariable region I, 291, 292*f*
    hypervariable region II, 291, 292*f*
    Oxford sequence, 291
Mitochondrial genome, 291*f*, 356*f*
Mitochondrial mutations, 355*f*, 356–357
Mitogen, 186
Mitosis, 181
Mixed chimerism, 285
Mixed leukocyte reaction, 429–430
Mixed lymphocyte culture (MLC), 418
MLC. *See* Mixed lymphocyte culture (MLC)
MLST. *See* Multilocus sequence typing (MLST)
MMR. *See* Mismatch repair (MMR)
Mobility, 139
Modified bases, 7*b*
Molecular abnormalities, 382
Molecular analysis
    affinity maturation, 391*b*
    immunoglobulin light-chain gene rearrangement in B cells, 389–390
    immunoglobulins heavy-chain gene clonality, 391–394, 392*f*, 393*f*, 394*f*
    MHC, 425–437
        combining typing results, 436, 436*t*
        coordination of HLA test methods, 437
        crossmatching, 427
        DNA-based typing, 430–436
        HLA test discrepancies, 436–437
        serological analysis, 425–437
    somatic hypermutation, 388
Molecular basis, inherited diseases, 345, 345*b*
Molecular Beacons, 161–162, 162*f*
Molecular biology, 77–257
    chromosomal structure and mutations, 179–198
    DNA sequencing, 223–257
    gene mutations, 199–222
    nucleic acid amplification, 142–178
    nucleic acid and protein analysis and characterization, 112–141
    nucleic acid extraction methods, 78–96
    nucleic acid resolution and detection, 97–111
Molecular detection, microorganism, 308–324
Molecular detection of resistance, 327–329
    antimicrobial resistance in *M. tuberculosis*, 329
    beta-lactam antibiotic resistance, 327–328
    glycopeptide antibiotic resistance, 328–329
    methicillin, 328
    minimum inhibitory concentration, 327
    susceptibility testing, 327
Molecular epidemiology, 329–337
    epidemic, 329
    molecular strain typing, 330–336
    pandemic, 329
    typing method comparison, 336–337, 337*t*
Molecular laboratory quality control, 305–307, 446–472. *See also* Laboratory techniques
    acrylic shielding, 466, 467*f*
    amplification control, 305, 457
    biohazard stickers, 447*f*
    block thermal cyclers, 461*f*
    cameras, gel photography, 462*f*
    centrifuge speeds, 461*f*
    certified chamber thermometers, 460*f*
    Clinical Laboratory Improvement Amendments, 447, 470
    collection tubes, nucleic acid testing, 449*f*, 449*t*
    contamination controls, 305
    cryotubes, 451, 453*f*
    documentation, test results, 468–469
    electrophoresis capillary replacement, 459*f*
    explosive materials, 450*f*, 465*f*
    flammable materials, 465*f*
    gel electrophoresis, 461*f*
    gloves for handling explosive materials, 450*f*
    hazard labels, 466*f*
    heterologous extrinsic, 305
    heterologous intrinsic, 305
    homologous extrinsic, 305
    instrument maintenance, 459–463
    internal controls, 305
    laminar flow hoods, 462*f*
    negative quality control, 457
    negative template control, 305
    nucleic acid storage, 453*t*
    positive controls, 305, 457
    radionuclides, 467*t*
    reagent blank, 305

Molecular laboratory quality control *(cont'd)*
  reagents, 463–468
  sensitivity control, 305, 457
  specimen handling, 447–451
  specimen storage, 452*t*
  test performance, 451, 453–458
  test performance measurements, 454*t*
  true negative, 305
  validation, 306
Molecular laboratory testing summary, 439–440
Molecular oncology, 369–416
  analytical targets, molecular testing, 372
    acute lymphocytic leukemia, 406*t*
    acute monocytic leukemia, 406*t*
    acute myeloid leukemia, 406*t*
    acute myeloid leukemia/myelodysplastic syndrome, 406*t*
    acute myelomonocytic leukemia, 406*t*
    acute nonlymphocytic leukemia, 406*t*
    acute promyelocytic leukemia, 406*t*
    acute T-lymphocytic leukemia, 406*t*
    anaplastic lymphoma receptor tyrosine kinase (*ALK*) proto-oncogene, 2p23.1, 382
    aneuploidy, 372
    ataxia telangiectasia mutated gene, *ATM* (11q22), 379–380
    B-cell leukemia, 406*t*
    breast cancer 1 gene, *BRCA1* (17q21), and breast cancer 2 gene, *BRCA2* (13q12), 380
    Burkitt lymphoma, 406*t*
    chronic lymphocytic leukemia, 406*t*
    chronic myelogenous leukemia, 406*t*
    clonality detection, 391–405
    diffuse large B-cell lymphoma, 406*t*
    epidermal growth factor receptor, *EGFR* (7p12), 373–375
    Ewing sarcoma, *EWS* (22q12), 376–377, 377*t*, 378*f*
    follicular lymphoma, 406*t*
    gene/chromosomal mutations, solid tumors, 372–387
    hairy cell leukemia, 406*t*
    Harvey rat sarcoma viral oncogene homolog, *H-ras* (11p15), 375–376, 375*f*, 375*t*
    hematological malignancies, molecular analysis, 387–406
    hematological malignancies, molecular targets, 387–397
    heterozygosity loss, 385
    human epidermal growth factor receptor 2, *HER2/neu/erb-b2 1* (17q21.1), 372–373
    Kirsten rat sarcoma viral oncogene homolog, K-ras (12p 12), 375–376, 375*f*, 375*t*, 376*f*
    leukemia and lymphoma molecular analysis, 388–391
    leukemias, chromosomal abnormalities, 406*t*
    liquid biopsy, 386–387
    lymphomas, chromosomal abnormalities, 406*t*
    mantle cell lymphoma, 406*t*
    microsatellite instability, 382, 384–385, 385*f*
    molecular abnormalities, 382
    mucosa-associated lymphoid tissue lymphoma, 406*t*
    multiple myeloma, 406*t*
    myeloproliferative/myeloblastic disease, 406*t*
    myeloproliferative/myelodysplastic disease, 406*t*
    neuroblastoma ras, *N-ras* (1p13), 375–376, 375*f*, 375*t*
    paired box-forkhead in rhabdomyosarcoma, *PAX3-FKHR*, *PAX7-FKHR*, t(1;13), t(2;13), 378
    polycythemia vera, 406*t*
    pre-B acute lymphoblastic leukemia, 406*t*
    proto-oncogenes, 381
    rearranged during transfection (*RET*) proto-oncogene (10q11), 381
    sequencing panels, 405
    synovial sarcoma translocation, chromosome 18-synovial sarcoma breakpoint 1 and 2, *SYT-SSX1*, *SYT-SSX2* t(X;18) (p11.2;q11.2), 377–378
    T-cell lymphoma, 406*t*
    T-chronic lymphocytic leukemia, 406*t*
    tissue-specific, 372
    tumor protein 53, *TP53* (17p13), 378–379
    tumor-specific, 372
    V-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog, *Kit*, *c-KIT* (4q12), 382
    V-myc avian myelocytomatosis viral-related oncogene, neuroblastoma-derived, 381
    Von Hippel-Lindau gene, *VHL* (3p26), 380–381
    V-Ros Avian UR2 Sarcoma Virus Oncogene Homolog 1 (*ROS1*) Proto-Oncogene (6q22.1), 381
    Waldenström macroglobinemia, 406*t*
  cancer molecular basis, 371–372
  capillary gel electrophoresis, 385*f*
  cell division cycle, 371*f*
  epidermal growth factor, 374*f*
  EWS translocation partners, 377*t*
  gene rearrangements. *See* Gene rearrangements
  germline, 387–388
  immunoglobulin, 389*f*
  immunoglobulin heavy-chain gene, 388*f*, 391–394, 392*f*, 393*f*
  immunoglobulin heavy-chain gene rearrangement, B cells, 387–389, 388*f*
  immunoglobulin light-chain kappa locus, 395*f*
  immunoglobulin light-chain lambda locus, 395*f*
  leukemias, chromosomal abnormalities, 406*t*
  microsatellite instability, 385*f*
  mismatch repair system, 383*t*, 384*f*
  neoplasm classification, 370–371
  p53 mutation analysis, 379*t*
  ras proteins, 377*f*
  receptor tyrosine kinases, 373*f*
  RT-PCR, 378*f*
  solid tumors, molecular abnormalities, 383*t*
  T-cell receptor gene, 390*f*, 390*t*, 395*f*, 396*f*
  V(D)J recombination, 387, 387*f*
Molecular strain typing, 330–336
  amplified fragment length polymorphism, 332–333
  arbitrarily primed PCR, 332
  coa typing, 335
  enterobacterial repetitive intergenic consensus, 333, 335*f*
  internal transcribed spacer elements, 334
  interspersed repetitive elements, 333–334
  ITS1, 334
  ITS2, 334
  locus-specific RFLP, 332
  mass spectrometry, 336
  multilocus sequence typing, 336, 336*t*
  PCR-RFLP, 332
  plasmid analysis, 331

  pulsed-field gel electrophoresis, 331
  REP-PCR, 333–334
  restriction fragment length polymorphism, 331–332
  ribotyping, 332
  spa typing, 334–335
Molecular-based detection methods, 310*t*, 312*t*
Monoclonal antibodies, 126
Monoclonal (monotypic) clonality, 391
Monomer, 43
Monomeric, 14
Morphology, chromosome, 183–184
Mosaic, 284*b*
Motif, 249*t*
mRNA. *See* Messenger RNA (mRNA)
MRSA. *See* Methicillin-resistant *S. Aureus* (MRSA)
MSI. *See* Microsatellite instability (MSI)
Mucosa-associated lymphoid tissue lymphoma, 406*t*
MUD. *See* Matched unrelated donor (MUD)
Mullis, Kary, 143, 145*b*, 146*b*
Multicolor FISH, 194, 194*f*
Multidrug-resistant *N. gonorrhoeae*, 326
Multifactorial inheritance, 363
Multilocus sequence typing (MLST), 336, 336*t*
Multimeric proteins, 347, 348*f*
Multiple displacement amplification, 166
Multiple endocrine neoplasia (MEN) syndromes, 381
Multiple locus probe, 265–266
Multiple myeloma, 406*t*
Multiple sequence alignment, 249*t*
Multiplex allele-specific PCR, 214*f*
Multiplex PCR, 156, 216, 216*f*, 286*f*
Mung bean nuclease, 18
Mutation
  chromosomal, 181, 186–195, 186*f*–195*f*
  copying errors resulting in, 12
  defined, 180
  detection methodologies, 208*t*
  gain-of-function, 347
  gene, 181, 199–222. *See also* Gene mutations
  genome, 181
  gonadal mosaicism, 348*b*
  hematological malignancies, 397–405
  lymphoid malignancies, 397–400
  in mitochondrial genes, 356–357
  myeloid malignancy, 401–405
  nuclear gene mutation disorders, 358*t*
  spectra, 405
Mycobacterium tuberculosis, 309–310
Mycology, 323–324
Mycoplasma pneumoniae, 310–311
Mycoplasma spp., 313
Myeloablative, 284, 287
Myeloid malignancy mutations, 401–405
Myeloproliferative/myeloblastic disease, 406*t*
Myeloproliferative/myelodysplastic disease, 406*t*

N

Nanney, David, 65
NARP mitochondrial point mutation, 358*f*
NASBA. *See* Nucleic acid sequence-based amplification (NASBA)
National DNA Database (NDNAD), 275
National DNA Index System (NDIS), 275
NDIS. *See* National DNA Index System (NDIS)
NDNAD. *See* National DNA Database (NDNAD)
Negative controls, 457
Negative template control, 153, 305
Neisseria gonorrhoeae, 311
Neoplasm, defined, 370
Neoplasm classification, 370–371
Nested PCR, 157–158, 158*f*
Neuroblastoma ras, *N-ras* (1p13), 375–376, 375*f*, 375*t*
New mutations, 348*b*
Next-generation sequencing, 238–248
  filtering and annotation, 247, 247*t*
  gene panels, 240
  Human Genome Project, 239*b*
  library, 239*f*, 435*f*
  library preparation, 239*f*, 240–241, 241*f*
  sequence quality, 246
  sequencing platforms, 243–246, 244*f*, 245*f*, 246*f*
  targeted libraries, 241–242, 242*f*, 243*f*
  test performance, 456
Nick, 13
Nick translation, 13, 13*f*, 126
NIRCA. *See* Nonisotopic RNase cleavage assay (NIRCA)
NIRCA analysis, 217*f*
Nirenberg, Marshall, 43, 44
Nitrocellulose, 117
Nitrogen bases, 3, 4
NMD. *See* Nonsense-mediated decay (NMD)
Noncoding RNAs, 69–73
Nonconservative substitution, 200
Nonconventional MHC antigens, 437
Nondisjunction, 346
Nonhistone proteins, 181, 183
Non-Hodgkin lymphoma, 370
Noninformative loci, 285
Nonisotopic RNase cleavage assay (NIRCA), 216, 217*f*
Nonpolar, 38
Nonpolyglutamine expansion disorders, 363*t*
Nonradioactive detection, 132*t*
Nonsense codons, 45
Nonsense substitution, 200
Nonsense-mediated decay (NMD), 51
Non-sequence-specific dyes, 160*f*
NOR. *See* Nucleolar organizing region (NOR) staining
Northern blot, 122, 133*f*
Nosocomial, 330
Nowell, Peter, 188*b*
NTPs, 150
Nuclear gene mutation disorders, 358*t*
Nuclease BaI31, 18
Nucleases, 17–18
Nucleated cells in suspension, 80
Nucleic acid, 8–9, 112–141
  amplification, tests of, 315*t*–317*t*
  amplification of, 142–178
  analyses, 207–218
  apurinic site, 119*f*
  biochemistry of, 2–37

Nucleic acid *(cont'd)*
  biotin, side chain, 127*f*
  capillary transfer, 120, 120*f*
  characterization, 112–141
  detection of, 97–111
  dioxetane substrates, light emitted, 130*f*
  dot blot, 134*f*
  electrophoretic transfer, 120, 121*f*
  extraction of, 78–96
  gel mobility shift assay, 139*f*
  genomic DNA fragments, 118*f*
  hybridization technologies, 116–121, 117*t*
  indirect nonradioactive detection, 130*f*
  nonradioactive detection, 132*t*
  northern blot, 133*f*
  peptide nucleic acids, 123, 125*f*
  phosphodiester, nucleic acids, 125*f*
  photolithographic target synthesis, 136*f*
  probe design of, 126–128
  probe types of, 124
  resolution of, 97–111
  restriction enzyme mapping, 113–115, 114*f*, 115*f*
  sample labeling, 137*f*
  sequence-based amplification of, 164–165
  single-stranded DNA, reannealing, 129*f*
  solution hybridization, 138*f*
  Southern blot, 133*f*
  storage of, 453*t*
  test, target sequences, 307*f*
  testing collection tubes, 449*f*, 449*t*
  vacuum transfer, 121, 121*f*
Nucleic acid sequence-based amplification (NASBA), 164–165
Nucleic acid-specific dyes, 110
Nuclein, 3
Nucleolar organizing region (NOR) staining, 186
Nucleolus, 28, 58
Nucleoside, 4, 5*f*
Nucleosome, 65, 181
Nucleotide deoxyguanosine 5′ phosphate, 5*f*
Nucleotide repeat expansion disorders, 357–362
  fragile X syndrome, 358–360, 359*f*, 360*f*
  Huntington disease, 361
  idiopathic congenital central hypoventilation syndrome, 361–362
  trinucleotide repeat, 362
  triplet-repeat expansion, 361, 362*f*
Nucleotide triphosphate, 10, 126
Nucleotides, 3, 4–8
Nucleotides deoxyguanosine monophosphate (dGMP), 6*f*
Null allele, 425

O

Okazaki fragments, 10
Oligo polythymine columns, 90*f*
Oligoclone, 391*b*
Oligomers, 43, 165*f*
Oligonucleotides, 18
Oncogenes, 371
Oncology, 369–416
  analytical targets, molecular testing, 372
  cancer molecular basis, 371–372
  capillary gel electrophoresis, 385*f*
  cell division cycle, 371*f*
  defined, 370
  epidermal growth factor, 374*f*
  EWS translocation partners, 377*t*
  gene rearrangements. *See* Gene rearrangements
  immunoglobulin, 389*f*
  immunoglobulin heavy-chain gene, 388*f*, 392*f*, 393*f*
  immunoglobulin light-chain kappa locus, 395*f*
  immunoglobulin light-chain lambda locus, 395*f*
  leukemias, chromosomal abnormalities, 406*t*
  microsatellite instability, 385*f*
  mismatch repair system, 383*t*, 384*f*
  neoplasm classification, 370–371
  p53 mutation analysis, 379*t*
  ras proteins, 377*f*
  receptor tyrosine kinases, 373*f*
  RT-PCR, 378*f*
  solid tumors, molecular abnormalities, 383*t*
  T-cell receptor gene, 390*f*, 390*t*, 395*f*, 396*f*
Oncomirs, 70
1000 Genomes Project, 253–254
Open reading frame, 31
Operator, 61
Operon, 61, 61*f*
Optimal alignment, 249*t*
Organic DNA isolation, 82*f*
Organic isolation methods, 81–83
Organic RNA extraction, 88, 88*f*
Orthology, 249*t*
Overhangs, 16
Oxacillin, 328*f*
Oxford sequence, 291

P

P site, 47, 49
p53 mutation analysis, 379*t*
PAGE. *See* Polyacrylamide gel electrophoresis (PAGE)
PAGE analysis, 285*f*
Paired box-forkhead in rhabdomyosarcoma, *PAX3-FKHR*, *PAX7-FKHR*, t(1;13), t(2;13), 378
Palindromic, 16
PAM. *See* Protospacer adjacent motif (PAM)
Pandemic, 329
Panel reactive antibodies (PRA), 429
Paracentric inversions, 188
Paralogy, 249*t*
Parasites, 324
Parentage testing, RFLP, 264–265, 265*f*
Partial dominance, 347
Paternity index, 275–276
Paternity tests, 276*t*
Pauling, Linus, 42
PCR. *See* Polymerase chain reaction (PCR)
Pedigree, 346, 347*f*
Penetrance, 349
Penicillin structure, 328*f*
Peptide, 38
Peptide bond, 38, 41*f*
Peptide nucleic acids (PNAs), 123, 125*f*

Peptidyl transferase, 49
Pericentric inversions, 188
Peristaltic pump, 106*f*
PFGE. *See* Pulsed-field gel electrophoresis (PFGE)
PFGE pattern interpretation, 332*t*
ph drift, 106
Phasing of alleles, 434
Phenotype, 180
Phosphodiester bond, 3, 125*f*
Phosphor, 134
Photobleaching, 193
Photographic equipment, 462
Photolithographic target synthesis, 136*f*
Phrap, 235*t*
Phred, 235*t*
Plasma, DNA isolation, 80–81
Plasma cell, 370
Plasmid analysis, 331
Plasmids, 23, 25–26, 26*f*
PNAs. *See* Peptide nucleic acids (PNAs)
Point mutations, 200, 200*t*
Polar, 38
Polarity, 8
Polonies, 243
PolyA polymerase, 36
PolyA tail, 31, 31*b*
Polyacrylamide
  electrophoresis of nucleic acids, 98
  repeating unit acrylamide, 101*f*
Polyacrylamide concentration, 103*t*
Polyacrylamide electrophoresis combs, 107*f*
Polyacrylamide gel electrophoresis (PAGE), 101–102
Polyadenylate polymerase, 31, 59
Polyadenylation, 31
Polyadenylation signal, 29
Polyadenylic acid, 31
Polycistronic RNA, 29
Polyclonal antibodies, 124–125
Polyclonal (polytypic) clonality, 391
Polyclonal lymphocytes, 391
Polycythemia vera, 406*t*
Polymerase, 28*f*, 58*t*
Polymerase chain reaction (PCR), 143–164
  amplicons, 143, 147*f*
  amplification control, 153
  amplification program, 144
  annealing, 145, 146*f*, 147*b*
  basic PCR procedure, 143–147
  buffer, 151
  complementary DNA, 31, 157
  components, 144*f*, 144*t*, 147–151
  contamination, 150*b*
  contamination control in, 153–154
  control, 152–156
  C_T, 160
  cycles, 144
  deaza dGTP, 149–150
  denaturation, 144
  deoxynucleotide triphosphates, 143, 146*f*
  deoxyribonucleotide bases, 149*b*
  DNA polymerase, 150–151
  DNA template, 149
  elements, 144*t*
  endpoint analysis, 159
  fidelity, 150
  fluorescent resonance energy transfer, 162–163, 163*f*
  Homogeneous MassExtend, 146*b*
  hot-start, 154–155
  Human Haplotype Mapping Project, 146*b*
  hybrid, 157
  mass spectrometry of viruses, 323
  master mix, 152
  mispriming, 148, 148*f*
  modifications, 156–163
  Molecular Beacons, 161–162, 162*f*
  multiplex, 156
  negative template control, 153
  nested PCR, 157–158, 158*f*
  NTPs, 150
  PCR-RFLP, 215–216, 215*f*, 332
  prevention of mispriming, 154
  primer dimers, 148, 149*f*
  primers, 147–150, 148*b*
  product cleanup, 155, 156*f*
  psoralens, 153
  pyrimidine dimers, 154*b*
  quantitative, 158–164, 159*f*
  quencher, 162
  rapid, 152
  reaction, 152, 152*f*
  reagent blank, 152
  real-time, 152, 158–164, 159*f*
  reverse transcriptase, 150, 157
  scorpion, 162, 163*f*
  semi-nested, 157–158
  sequence-specific, 156–157
  spin columns, 155, 156*f*
  Stoffel fragment, 150
  strand, 143
  SYBR green, 160
  tailed primers, 149, 149*f*
  Taq polymerase, 150
  TaqMan, 160, 161*f*
  template, 143
  thermal cyclers, 151–152
  threshold cycle, 160
  touchdown, 155
  Tth polymerase, 150
  viruses, 149
Polymerase classified by sequence homology, 11*t*
Polymerases, 11–14, 35–36
Polymorphisms, 180, 262*t*
Polypeptides, 38, 420*f*
Polyphred, 235*t*
Polyploidy, 345
Polythymine, 31
Position effect, 67*b*, 181
Positive controls, 305, 457
Post-PCR, 153
Power supplies, 461

PRA. *See* Panel reactive antibodies (PRA)
Preanalytical error, 447
Pre-B acute lymphoblastic leukemia, 406*t*
Pre-PCR, 153
Precautions in sample processing, 450
Precision, 454*t*
Prehybridization, 121–122
Premutations, 359
Preparation of resolved DNA for blotting, 118–119
Preparing the DNA sample, 79–81
Prevention of mispriming, 154
Primary antibody, 123
Primary resistance mutations, 319
Primary structure, 41
Primase, 10
Primer dimers, 148, 149*f*
Primer PCR. *See* Sequence-specific PCR (SSP-PCR)
Primers
  DNA replication, 11*b*
  nucleic acid amplification, 147–150, 148*b*, 148*f*, 149*f*
  reagents, 463–464, 463*f*
  TaqMan probes, 161*f*
Prions, 43
Prior odds, 276–277, 277*t*
Private antibodies, 428
Probability of paternity, 276–277
Proband, 385
Probe, 117
Probe amplification, 168–171
  ligase chain reaction, 168–169, 169*f*
  Qβ replicase, 170–171, 171*f*
  strand displacement amplification, 169–170, 170*f*, 171*f*
Probes, 123–128
  DNA probes, 123–124
  locked nucleic acids, 123
  nucleic acid probe design, 126–128
  nucleic acid probe types, 124
  peptide nucleic acids, 123, 125*f*
  probe labeling, 126
  protein probes, 124–126
  RNA probes, 124
  secondary antibody, 123
Probit analysis, 455
Processivity, 14
Product rule, 274
Proficiency testing, 468
Profile, human identification, 266
Profile matching, 273–275
  Hardy-Weinberg equilibrium, 274*b*
  likelihood ratio, 274
  linkage disequilibrium, 282
  linkage equilibrium, 274
  match probability, 273
  product rule, 274
  subpopulations, 274
Prokaryotes, 30*f*, 62*f*, 63*f*
Prokaryotic RNA polymerase, 35*f*
Promoter, 28, 58
Proofread, 13
Prosthetic group, 43
Protease, 24–25
Protein analysis, 112–141
  apurinic site, 119*f*
  array-based hybridization, 133–138
  biotin, side chain, 127*f*
  capillary transfer, 120, 120*f*
  detection systems, 129–132
    nonradioactive detection, 130, 131*f*
    probe labeled with radioactive phosphorous atoms, 130, 130*f*
    signal-to-noise ratio, 132
  dioxetane substrates, light emitted, 130*f*
  dot blot, 134*f*
  electrophoretic transfer, 120, 121*f*
  gel mobility shift assay, 139*f*
  genomic DNA fragments, 118*f*
  hybridization conditions, stringency, 128–129
  hybridization technologies, 116–121, 117*t*
    northern blot, 122
    Southern blot, 117–121
    western blot, 123–124
  indirect nonradioactive detection, 130*f*
  ink-jet technology, 135*f*
  interpretation of results, 132–133
  nonradioactive detection, 132*t*
  northern blot, 133*f*
  peptide nucleic acids, 123, 125*f*
  phosphodiester, nucleic acids, 125*f*
  photolithographic target synthesis, 136*f*
  probes, 123–128
  restriction enzyme mapping, 113–115, 114*f*, 115*f*
  sample labeling, 137*f*
  single-stranded DNA, reannealing, 129*f*
  solution hybridization, 138–139, 138*f*
  Southern blot, 133*f*
  vacuum transfer, 121, 121*f*
  western blot, 122–123, 123*b*, 127*b*, 127*f*, 133*f*
Protein gel electrophoresis, 430
Protein probes, 124–126
  alternative nucleic acids, 126*b*
  haptens, 125
  hybridomas, 126
  monoclonal antibodies, 126
  polyclonal antibodies, 124–125
Protein synthesis, 47–51, 50*f*
Protein truncation tests, 380
Protein-based identification, 293–294
Proteins, 37–46
  A site, 49, 49*f*
  alpha helix, 42, 42*f*
  amino acids, 38–43, 39*f*, 40*t*
  amino acyl tRNA synthetases, 47*f*
  amino terminal, 38
  aminoacyl tRNA synthetases, 46
  anticodon, 45–47
  beta-pleated sheets, 42
  carboxy terminal, 38
  chaperones, 49
  chromosome, 43
  codons, 45, 45*f*
  complement, 44*b*

  - conjugated proteins, 43
  - consensus, 42*b*
  - defined, 37
  - domains, 38
  - E site, 49
  - ER stress, 51*b*
  - extracellular domain, 38
  - gene, 43, 43*f*
  - genetic code, 38, 43–44, 45*f*
  - glycoproteins, 43
  - isodecoders, 46
  - leucine zipper, 42*b*
  - lipoproteins, 43
  - metalloproteins, 43
  - monomer, 43
  - nonpolar, 38
  - nonsense codons, 45
  - nonsense-mediated decay, 51
  - oligomers, 43
  - P site, 47, 49
  - peptide, 38
  - peptide bond, 38, 41*f*
  - peptidyl transferase, 49
  - polar, 38
  - polypeptides, 38
  - primary structure, 41
  - prions, 43
  - prosthetic group, 43
  - protein synthesis, 50*f*
  - proteome, 38
  - proximal regulatory sequences, 43
  - pyrrolysine, 38*b*
  - quaternary structure, 43
  - random coils, 42
  - regulatory sequences, 43
  - release factors, 51
  - ribosomal binding site, 47
  - ribosomes, 47, 48*f*
  - RNA Tie Club, 44*b*
  - secondary structure, 42
  - selenocysteine, 46*b*
  - selenoproteins, 46*b*
  - side chain, 38
  - structural sequences, 43
  - suppressor tRNAs, 47*b*
  - tertiary structure, 42–43
  - translation, 46–51
    - amino acid charging, 46–47
    - protein synthesis, 47–51, 50*f*
  - transmembrane proteins, 40*f*
  - tRNA charging, 46
  - zinc finger, 42*b*
  - zwitterions, 38
- Proteolytic lysis
  - DNA extraction/storage cards, 86
  - extraction with chelating resin, 86
  - of fixed material, 85–86
  - rapid extraction methods, 86
  - rapid lysis methods, 86
- Proteome, 38, 136*b*
- Prothrombin, 349–350, 351*t*, 352
- Proto-oncogenes, 381
- Protospacer, 17*b*
- Protospacer adjacent motif (PAM), 17*b*
- Proximal regulatory sequences, 43
- Pseudogenes, 261
- Psoralens, 153
- Public antigens, 428
- PubMed, 249*t*
- Pulsed-field gel electrophoresis (PFGE), 100–101, 331, 332*t*
- Purines, 4
- Pyrimidine dimers, 14*b*, 154*b*
- Pyrimidines, 4
- Pyrogram, 236
- Pyrophosphate, 32*b*
- Pyrophosphate exchange, 12
- Pyrophosphorolysis, 12
- Pyrosequencing, 235–236
  - defined, 236*f*
  - pyrogram, 236
- Pyrrolysine, 38*b*

Q

- Q banding, 184–185, 185*f*
- Qβ replicase, 170–171, 171*f*
- Quality assurance
  - molecular laboratory quality control, 458–459
  - surgical sections using STR, 289
  - testing, 289*f*
- Quality control. *See* Molecular laboratory quality control
- Quantitative FISH, 195*b*, 195*f*
- Quantitative PCR, 158–164, 159*f*
- Quaternary structure, 43
- Quencher, 160, 162
- Query, 248, 249*t*

R

- R banding, 185*b*
- R factors, 26
- Radioactive chemicals, 465, 467*f*
- Radioactive protein, 24*f*
- Radionuclides, 467*t*
- Random coils, 42
- Random priming, 126
- Randomly amplified polymorphic DNA, 167
- RAPD gel results, 333*f*
- Rapid extraction methods, 86
- Rapid lysis methods, 86
- Rapid PCR, 152
- Ras proteins, 377*f*
- Read alignment, 246
- Reagent alcohol, 468
- Reagent blank, 152
- Reagent blank controls, 305
- Reagents, 463–468
  - analyte-specific reagents, 464
  - categories, 464–465
  - chemical safety, 465–466
  - investigational use only, 465
  - labeling, 466–468

Reagents *(cont'd)*
  proficiency testing, 468
  reference range, 465
  research use only, 465
  storage, 466
  in vitro analytical test, 465
  in vitro diagnostic, 464–465
Real-time PCR, 152, 158–164, 159*f*
Reannealing, single-stranded DNA, 129*f*
Rearranged during transfection proto-oncogene (10q11), 381
recBC nuclease, 18
Receptor tyrosine kinases, 372, 373*f*
Recessive, 438
Reciprocal translocations, 186
Recognition sites, 8*b*
Recombinant DNA, 19, 21
Recombination, 22*f*
Recombination activating genes, 388
Recombination DNA, 19, 21*f*, 22*f*
Recombination in asexual reproduction, 21–25
  conjugation, 21–23
  transduction, 23
  transformation, 23–25
Recombination in sexually reproducing organisms, 19–21
Recombined chromosomes, 22*f*
Redundant tilting, 212
Reference range of tests, 454*t*, 465
Refrigerators and freezers, 459–460
REF-SSCP. *See* Restriction endonuclease fingerprinting (REF-SSCP)
Regulation of transcription, 60–65
  epigenetics, 65–69
  regulation of RNA synthesis at initiation, 60–64
Regulatory sequences, 43
Release factors, 51
Repairing polymerases, 14*b*
Replication, 9–14
Replication complex, 11*b*
Replication factor (RF), 226
Replication fork, 10
Replisome, 11*b*
Reportable range of tests, 454*t*, 455, 463
Reporter dye, 174
Reporting results, 469–470
REP-PCR, 333–334
Repressor, 61
Reproducibility of test results, 454*t*
Reproducible typing method, 337
Research use only (RUO), 465
Resequencing, 469
Resistance mechanisms, 326*t*
Resistance to antimicrobial agents, 325–327
  multidrug-resistant *N. gonorrhoeae*, 326
  multidrug-resistant *S. aureus*, 326
  transposon, 327
Resolution
  allele detail, 425
  HLA typing methods, 436*t*
  nucleic acid, 97–111
    agarobiose, 99*f*
    agarose, 99*t*
    capillary electrophoresis, 103*f*
    double-stranded DNA, 100*f*
    double-stranded DNA fragments, 102*f*
    dye comigration, 109*t*
    field inversion gel electrophoresis, 100*f*
    horizontal gel electrophoresis, 98*f*
    horizontal submarine gel, 106*f*
    peristaltic pump, 106*f*
    polyacrylamide, 101*f*
    polyacrylamide concentration, 103*t*
    polyacrylamide electrophoresis combs, 107*f*
    vertical gel apparatus, 108*f*
Respiratory tract pathogens, 309–311
  Bordetella pertussis, 309
  Chlamydophila pneumoniae, 311
  Legionella pneumophila, 309
  Mycobacterium tuberculosis, 309–310
  Mycoplasma pneumoniae, 310–311
  Streptococcus pneumoniae, 311
  typical organisms, 310*t*
Respiratory viruses, 322–323
Restriction endonuclease fingerprinting (REF-SSCP), 209*b*
Restriction enzyme cutting and resolution, 117–118
Restriction enzyme mapping, 113–115
Restriction enzymes, 14–17, 15*t*, 16*f*
Restriction fragment length polymorphism (RFLP)
  *Bam*H1, 215*f*
  defined, 115
  inheritance, 264*f*
  molecular epidemiology, 331–332
  multiplex PCR, 216*f*
  nucleic acid analyses, 215–216
  single-nucleotide polymorphisms, 261
  typing, 262–266
    alleles, 263
    genetic mapping with RFLPs, 264
    heterozygous, 263
    homozygous, 263
    human identification, RFLP, 265–266
    locus, 263
    parentage testing, RFLP, 264–265, 265*f*
Restriction map, 113
Restriction mapping, 114*f*, 115*f*
Restriction modification, 7*b*
Reverse dot blot, 134, 209
Reverse dot-blot SSOP, 431–432, 432*f*
Reverse transcriptase, 150
Reverse transcriptase PCR, 150, 157, 302
Reverse transcription, 27*b*
RF. *See* Replication factor (RF)
RFLP. *See* Restriction fragment length polymorphism (RFLP)
Rhabdomyosarcoma, 378
rho, 29
Ribonucleases, 24–25, 36–37
Ribonucleic acid. *See* RNA
Ribose, 27
Ribosomal binding site, 47
Ribosomal RNA (rRNA), 29, 30*f*, 336*f*
Ribosomes, 33, 47, 48*f*
Ribotyping, 332

Ring chromosome, 188
RISC. *See* RNA-induced silencing complex (RISC)
RNA, 27–37
  activator, 63
  alpha phosphate, 29, 59
  alternative splicing, 33
  anticodon, 34
  antisense strand, 59
  attenuation, 63, 63*f*
  base, 27
  base pairs, 70
  cap, 31, 32*b*
  chaperones, 49, 50*f*
  chromatin, 18, 65
  cis factors, 60, 61*f*
  constitutive, 30, 60
  corepressor, 63
  enhancers, 63
  enzyme adaptation (induction), 61, 61*b*
  enzyme induction, 63
  enzyme repression, 63
  epigenetic regulation, 66
  exons, 32
  gene expression, 27, 58
  hairpin, 29, 59
  histone, 65–66
  IncRNA, 71–72, 73*f*
  inducible transcription, 30
  intervening sequences, 31*b*
  introns, 32
  isolation, 87–90
    diethyl pyrocarbonate, 87*b*
    extraction of total RNA, 87
    intracellular, 88
    microfuge, 88–89
    polyA RNA, 89–90
    proteolytic lysis of fixed material, 89
    RNAse-free, 87
    specimen collection, 87
    total RNA, 87
  long noncoding, 71–72
  messenger RNA, 27, 58
  methylation, 69
  microRNAs, 70
  molecular techniques for, 3
  noncoding, 69–73
  nucleolus, 28
  nucleosome, 65
  open reading frame, 31
  operator, 61
  operon, 61, 61*f*
  polyA tail, 31
  polyadenylate polymerase, 59
  polyadenylation signal, 29
  polyadenylic acid, 31
  polycistronic, 29
  polymerases, 27, 28*t*, 35–36, 35*b*, 58, 58*t*
  polythymine, 31
  primers, 11*b*
  promoter, 28, 58
  regulation of transcription, 60–65
  repressor, 61
  reverse transcription, 27*b*
  rho, 29
  ribose, 27
  ribosomal RNA, 29, 30*f*, 336*f*
  ribosomes, 33, 47, 48*f*
  RNA interference, 70
  RNA-induced silencing complex, 71
  RNA-metabolizing enzymes, 36–37
  sense strand, 59
  silencers, 63
  Sm proteins, 34*b*
  splicesosome, 34*b*
  splicing, 31–33
  trans factors, 60
  transcription, 27–29
  transfer RNA, 33–34, 35*b*
  types/structures of RNA, 29–35
    messenger RNA, 29–33
    micro RNAs, 70
    ribosomal RNA, 29, 30*f*
    small interfering RNA, 70
    small nuclear RNA, 33, 33*t*
    small RNAs, 70–71
    transfer RNA, 33–34, 35*b*
  uracil, 27
RNA editing, 36*b*
RNA integrity control, 401
RNA interference (RNAi), 70
RNA isolation, specimen sources, 90*t*
RNA polymerases, 27, 28*t,* 35–36, 35*b*, 58, 58*t*
RNA probes, 124
RNA sequencing, 237–238
RNA Tie Club, 44*b*
RNA-dependent RNA polymerases, 35
RNAi. *See* RNA interference (RNAi)
RNA-induced silencing complex (RISC), 71
RNA-metabolizing enzymes, 36–37, 37*t*
RNase protection, 138
RNases, 37*t*
RNases H, 11*b*
RNA-SSCP, 209*b*
Robertsonian translocation, 187, 189*f*
rRNA. *See* Ribosomal RNA (rRNA)
rSSCP, 209*b*
RT-PCR, 378*f*
RUO. *See* Research use only (RUO)

S

S1 nuclease, 18
Safety Data Sheets, 467
Salting out, 83
Sample labeling, 137*f*
Sample preparation, 304–305
Sanger sequencing. *See* Dideoxy chain termination (Sanger) sequencing
Sarcomas, 370
Satellite plasmid DNA, 26
Scorpion primer/probes, 162, 163*f*
Screening, 428–429

SDA. *See* Strand displacement amplification (SDA)
Secondary antibody, 123
Secondary resistance mutations, 319
Secondary structure, 42
Selenocysteine, 38*b*
Selenoproteins, 38*b*
Self-sustaining sequence replication (SSSR), 164–165
Self-transmissible plasmid, 26
Semiconservative, 9
Semi-nested PCR, 157–158
Sense strand, 59
Sensitivity control, 305, 457
SeqScape, 235*t*
Sequence complexity, 128
Sequence interpretation
  base calling, 232
  dye blobs, 232, 233*f*
  mutations or polymorphisms, 234*f*
  sequence quality, 233*f*
  software programs, 232, 235*t*
Sequence of nucleotides, 4, 180
Sequence targets, 307
Sequence-based typing, 433–435, 433*f*, 434*f*
Sequence-specific oligonucleotide probe hybridization (SSOP), 431–432, 431*f*
Sequence-specific PCR (Sequence-specific primer PCR, SSP-PCR)
  factor V Leiden, 352*f*
  molecular analysis of the MHC, 432–433, 432*f*, 433*f*
  nucleic acid analyses, 213–214
  target amplification, 156–157
Sequence-specific primer amplification, 213*f*
Sequencing DNA, 223–257
  Assign, 235*t*
  bioinformatics, 248–250
  bioinformatics terminology, 249*t*
  bisulfite DNA sequencing, 236–237
  BLAST, 235*t*
  consensus sequences, 248
  cytosine residues, 238*f*
  dideoxynucleotide, 226*f*
  direct sequencing, 224–235
  electropherogram, 233*f*
  Factura, 235*t*
  FASTA, 235*t*
  fluorescent sequencing chemistries, 231*f*
  GRAIL, 235*t*
  Human Genome Project, 250–254, 250*t*, 251*t*, 253*t*
  IUB universal nomenclature, 248
    Basic Local Alignment Search Tool, 248
    hierarchal shotgun sequencing, 251, 252*f*
    for mixed bases, 250*t*
  manual dideoxy sequencing, 225*f*
  Matchmaker, 235*t*
  Maxam-Gilbert sequencing, 224*t*, 225*f*
  next-generation sequencing, 238–248, 239*b*, 239*f*, 456
  Phrap, 235*t*
  Phred, 235*t*
  Polyphred, 235*t*
  pyrosequencing, 235–236, 236*f*
  SeqScape, 235*t*
  sequencing ladder, 230
  software programs, sequence data, 235*t*
  TIGR Assembler, 235*t*
Sequencing ladder, 230
  DNA synthesis, 228*f*
  reading of, 226*f*
  sequencing, RNA, 237–238
Sequencing library, 239–240
Sequencing panels, 405
Sequencing-based methods, 213–215
  allelic discrimination with fluorogenic probes, 214–215, 215*f*
  sequence-specific PCR, 213–214
Serological analysis, 427–430
  alloantibodies, 428
  antihuman antibodies, 428
  complement-dependent cytotoxicity, 427
  crossmatching, 429, 430*f*
  cross-reactive epitope groups, 428
  HLA typing, 427–428
  humoral sensitization, 428
  mixed leukocyte reaction, 429–430
  panel reactive antibodies, 429
  private antibodies, 428
  protein gel electrophoresis, 430
  public antigens, 428
  screening, 428–429
  typing trays, 427
Serologically defined HLA specificities, 422*t*–424*t*
SETS. *See* Swab Extraction Tube System (SETS)
Sex-linked disorders, 348
Sexually reproducing organisms, recombination in, 19–21
Shark tooth combs, 107
Short interspersed nucleotide sequence (SINES), 261
Short tandem repeat TH01, 267*f*
Short tandem repeats (STR), 261, 267, 268*f*. *See also* STR *subjects*
Sibling index, 277
Sibling tests, 277–278
Side chain, 38
Sigma factor, 35*b*
Signal amplification, 172–174
  branched DNA amplification, 172–173, 172*f*
  cleavage-based amplification, 173–174, 174*f*
  cycling probe, 174
  hybrid capture assays, 173, 173*f*
Signal-to-noise ratio, 132
Signal transduction, 375*b*
Silencers, 63
Silent nucleotide substitution, 200
Silver-enhanced in situ hybridization (SISH), 373
Silver stain, 90, 110–111
Similarity, 249*t*
Simultaneous replication, 10*f*
SINES. *See* Short interspersed nucleotide sequence (SINES)
Single nucleotide polymorphisms (SNPs), 289–291
  defined, 261
  Human Haplotype Mapping project, 290–291, 290*f*, 291*b*
  types, 290*t*
Single-gene disorders, 355–363
  inheritance patterns, 346–349
  molecular basis, 349–355

  molecular methods, 350*t*
  nonclassical patterns, 355–363
Single-locus probe (SLP), 266, 266*f*
Single-strand break, 19
Single-strand conformation polymorphism (SSCP), 207–209, 208*f*
Single-stranded DNA, reannealing, 129*f*
SISH. *See* Silver-enhanced in situ hybridization (SISH)
Sizes of human chromosomes, 180*t*
Slot blot, 133–134, 134*f*
SLP. *See* Single-locus probe (SLP)
Sm proteins, 34*b*
Small interfering RNA, 70
Small nuclear RNA, 33, 33*t*
Small RNAs, 70–71
Snell, G., 418*b*
SNPs. *See* Single nucleotide polymorphisms (SNPs)
Software programs, 235*t*
Solid matrix RNA isolation, 89*f*
Solid tumors, molecular abnormalities, 383*t*
Solid-phase emulsion PCR, 166–167, 167*f*
Solid-phase isolation, 84–85, 84*f*
Solution hybridization, 138–139, 138*f*
  gel mobility shift assay, 139
  RNase protection, 138
  solution hybridization, 138–139
Somatic hypermutation, 388
Southern blot
  defined, 117
  example, 133*f*
  nitrocellulose, 117
  preparation of resolved DNA for blotting, 118–119
  probe, 117
  restriction enzyme cutting and resolution, 117–118
  single nucleotide polymorphisms, 262*f*
Spa typing, 334–335
Specimen collection, 302–304, 448–450
Specimen handling, 447–451
  collection tubes for molecular testing, 448–450
  contact precautions, 450
  criteria for accepting specimens, 447–448, 448*f*
  cryotubes, 451, 453*f*
  holding and storage requirements, 451
  preanalytical error, 447
  precautions in sample processing, 450
  standard precautions, 450
  transmission-based precautions, 450
Specimen storage, 452*t*
Specimen transport systems, 303*t*
Spectral karyotyping, 193
Spectrophotometry, 92–93
Spin columns, 155, 156*f*
Splicesosome, 34*b*
Splicing, 31–33, 32*f*, 33*b*
Split chimerism, 286*b*
Split specificities, 422
SSCP. *See* Single-strand conformation polymorphism (SSCP)
SSOP. *See* Sequence-specific oligonucleotide probe hybridization (SSOP)
SSOP assay, 431*f*
SSP-PCR. *See* Sequence-specific PCR (SSP-PCR)
SSSR. *See* Self-sustaining sequence replication (SSSR)
Staggered separation of the duplex, 16
Staining pattern, chromosomes, 185*f*
Standard curve, 457
Standard precautions, 450
Standard tiling, 212, 212*f*
Star activity, 114
Sticky ends, 16
Stoffel fragment, 150
Stop codon, 200
Storage diseases, 351*t*
Storage of reagents, 466
STR. *See* Short tandem repeat (STR)
STR analysis, 268–278
STR locus information, 271*t*
STR nomenclature, 270
STR TH01, 267*f*
STR typing by PCR, 266–278
  allelic ladders, 268
  discrete allele systems, 268
  gender identification, 270, 270*f*
  internal size standards, 268
  microsatellites, 267
  microvariants, 267
  mini-STR, 268
  short tandem repeats, 267
  STR analysis, 268–278
  STR nomenclature, 270
  test results analysis, 270, 272, 272*b*, 272*t*, 273*b*
Strand, 143
Strand displacement amplification (SDA), 169–170, 170*f*, 171*f*
Streptococcus pneumoniae, 311
Stringency, 128
Structural sequences, 43
Stutter, 286–287
Submarine gels, 106
Submetacentric chromosome, 184, 184*f*
Substituted nucleotides, 7–8, 7*f*
Substrate specificity, 14
Subtelocentric chromosomes, 184*f*
Sugar-phosphate backbone, 8
Supercoiled DNA, 19
Supercoiled plasmids, 26, 26*f*
Suppressor tRNAs, 47*b*
Surface amplification, 167, 168*f*
Surname test, 282
Susceptibility testing, 327
Swab Extraction Tube System (SETS), 303*f*
SwissProt, 249*t*
SYBR green, 160
Syngeneic transplant, 284*b*
Synonymous changes, 425
Synovial sarcoma translocation, 377–378

T

T: IUB universal nomenclature, 248
  Basic Local Alignment Search Tool, 248
  hierarchal shotgun sequencing, 251, 252*f*
t(1;13), t(2;13), 378
t(8;14)(q24;q11), 400, 400*f*

t(9;22)(q34;q11), 401–403, 401*f*, 402*f*, 403*f*
t(11;14)(q13;q32), 398–399, 400*f*
t(14;18)(q32;q21), 397–398, 397*f*, 398*f*, 399*f*
t(15;17)(q22;q11.2-q12), 403–404, 404*f*
Tailed primers, 149, 149*f*
Tandem repeat, 265*f*
Taq polymerase, 150
TaqMan, 160, 161*f*
TaqMan probe, 162*f*
TaqMan signal fluorescence, 162*f*
Target, 143
Target amplification, 143–168
  genomic amplification methods, 165–168
  polymerase chain reaction, 143–164
  transcription-based amplification systems, 164–165
Target sequences, nucleic acid test, 307*f*
Targeted libraries, 241–242, 242*f*, 243*f*
Tatum, Edward L., 23*b*
T-cell lymphoma, 406*t*
T-cell receptor gene, 395*f*, 396*f*
T-cell receptor gene rearrangements, 390–391, 390*f*, 390*t*, 391*f*, 394–396, 395*f*, 396*f*
T-chronic lymphocytic leukemia, 406*t*
TE buffer, 83*b*
Telocentric chromosome, 184
Telomere, 192
Telomeric probe, 192, 193*f*, 194*f*
TEMED (tetramethylethylenediamine), 102
Template, 9, 143, 224
Teratocarcinomas, 370
Terminal transferase, 14
Termination, transcription, 29
Tertiary structure, 42–43
Test
  allelic frequencies, paternity testing, 275–277
  analytical targets, molecular testing, 372
  avuncular, 277–278
  bone marrow engraftment, 283–289
  calibrators, 456–457
  controls, 457–458
  documentation of test results, 468–469
  FDA approved, 455
  FDA cleared, 455
  laboratory-developed, 451
  next-generation sequencing, 456
  nucleic acid amplification, 315*t*–317*t*
  parentage, 264–265, 265*f*
  paternity, 276*t*
  performance. *See* Test performance
  proficiency, 468
  quality assurance, 289*f*
  sibling, 277–278
  surname, 282
  validation, 455
  in vitro analytical test, 465
Test performance, 451, 453–458
  amplification control, 457
  analyte measurement range, 454*t*
  analytic accuracy, 454, 454*t*
  analytic sensitivity, 454, 454*t*
  analytic specificity, 454, 454*t*
  calibration curve, 457
  clinical sensitivity, 454*t*
  clinical specificity, 454*t*
  controls, 457–458
  cut-off values, 458
  detection limit, 454*t*, 455
  high-positive controls, 457
  internal controls, 457
  linearity, 454*t*
  low-positive controls, 457
  measurements, 454*t*
  negative controls, 457
  positive controls, 457
  precision, 454*t*
  quality assurance, 458–459
  reference range, 454*t*
  reportable range, 454*t*
  reproducibility, 454*t*
  sensitivity control, 457
  standard curve, 457
Tetramethylethylenediamine (TEMED), 102
Thermal cyclers, 151–152
Thermometers, certified chamber, 460*f*
Threshold cycle, 160
Thrombosis, 351*t*
Thymidine, 4
Thymine, 3, 14*b*
Thymine dimers, 14*b*
TIGR Assembler, 235*t*
Tiselius, Arne, 98*b*
Tissue fixatives, 81*t*
Tissue samples, 81
Tissue typing, 417–445, 418
  additional recognition factors, 437–438
  allografts, 418
  bead arrays, serum antibody detection, 429*f*
  CDC assay example, 428*t*
  CDC expression, 428*t*
  crossmatching, antibodies, 427*f*
  cytotoxicity, cells stained for, 428*f*
  DNA polymorphisms, 421*f*
  graft-versus-host disease, 418
  HLA polymorphisms, 420–425
    allele, 420, 421–422
    haplotype, 420, 421*f*
    HLA nomenclature, 420, 422–425
    HLA type, 420
  KIR gene cluster, 438*f*
  major histocompatibility complex, 418, 419*t*
  MHC disease association, 438–439
  MHC locus, 418–420
  molecular analysis of MHC, 425–437
    combining typing results, 436, 436*t*
    coordination of HLA test methods, 437
    crossmatching, 427
    DNA-based typing, 430–436
    HLA test discrepancies, 436–437
    serological analysis, 425–437
  polypeptides, 420*f*

  reverse dot-blot SSOP, 431–432, 432*f*
  SSOP assay, 431*f*
  transplant evaluation, 439*t*
Tissue-specific molecular testing, 372
TMA. *See* Transcription-mediated amplification (TMA)
Topoisomerase inhibitors, 19
Topoisomerases, 19
Touchdown PCR, 155
tracrRNA, 115
Tracking dye, 108–109
Traits, 3
Trans factors, 60
Transcription, 27–29
  defined, 27, 58
  elongation, 28–29, 28*f*, 58–59
  factors, 64*f*
  initiation, 28, 58
  post-transcriptional regulation, 64
  regulation of epigenetics, 65–69
  regulation of RNA synthesis at initiation, 60–64
  termination, 29, 59–60
Transcription-based amplification systems, 164–165
  nucleic acid sequence-based amplification, 164–165
  self-sustaining sequence replication, 164–165
  steps, 164*f*
  transcription-mediated amplification, 164–165
Transcription-mediated amplification (TMA), 164–165
Transcriptome, 136*b*
Transduction, 23
Transfer RNA (tRNA), 33–34, 35*b*, 46
Transformation, 23–25
Transforming factor, 24*f*, 25
Translation, 46–51
  amino acid charging, 46–47
  post-translational regulation, 64–65
  protein synthesis, 47–51, 50*f*
Translocations, 186
Transmembrane proteins, 40*f*
Transmission patterns, 346
Transmission-based precautions, 450
Transplant evaluation, 439*t*
Transposon, 327
Treponema pallidum, 311–313
Trimming of nucleotides, 391
Trinucleotide-repeat disorders, 362
Triploid, 346
Trisomy, 181
tRNA. *See* Transfer RNA (tRNA)
tRNA charging, 46
True negative, 305
Tth polymerase, 150
Tumor protein 53, *TP53* (17p13), 378–379
Tumor-specific molecular testing, 372
Tumor-suppressor genes, 371
t(X;18)(p11.2;q11.2), 377–378
Type I restriction enzymes, 15
Type II restriction enzymes, 16, 16*b*, 16*f*
Type III restriction enzymes, 15
Types of gene mutations, 200–201
Types of polymorphisms, 261
Types/structures of RNA, 29–35
  small interfering RNA, 70
  small nuclear RNA, 33
  small RNAs, 70–71
  transfer RNA, 33–34, 35*b*
Typing capacity, 336
Typing method comparison, 336–337, 337*t*
  discriminatory power, 337
  reproducible, 337
  typing capacity, 336–337
Typing trays, 427
Tyrosine-kinase activity, 372, 373*f*
Tyrosine-kinase inhibitors, 374

U
Unbalanced, 187
Ultraviolet light (uv), 462, 462*f*
Uniparental disomy, 362–363
Uracil, 27, 27*f*
Urogenital tract pathogens, 311–313
  Chlamydia trachomatis, 311
  Mycoplasma spp., 313
  Neisseria gonorrhoeae, 311
  Treponema pallidum, 311–313

V
Vacuum transfer, 121, 121*f*
Validation, 306
Vancomycin structure, 329*f*
Vancomycin-resistant *S. aureus*, 327*f*
Variable expressivity, 349
Variable-number tandem repeat (VNTR), 261
Variant, 180
Varicella zoster virus, 321
V(D)J recombination, 387, 387*f*
Vertical gel apparatus, 108*f*
Vertical gel electrophoresis, 98*f*
VHL. *See* Von Hippel-Lindau syndrome (VHL)
Viral load, 317
Viral load measurement, test performance, 318*t*
Viruses, 313–323
  BK/JC viruses, 323
  cytopathic effect, 313–314
  DNA isolation, 80
  DNA template, 149
  hepatitis C virus, 321–322
  herpes viruses, 320–321
  human immunodeficiency virus, 314, 317–319
  human papillomavirus, 322
  mass spectrometry, 323
  mycology, 323–324
  nucleic acid amplification tests, 315*t*–317*t*
  respiratory viruses, 322–323
Visualizing chromosomes, 184–186
  4′,6-diamidino-2-phenylindole, 186
  C banding, 185*b*
  G bands, 185, 185*b*, 186*b*
  high-resolution banding, 186
  nucleolar organizing region staining, 186

Visualizing chromosomes *(cont'd)*
  Q banding, 184–185, 185*f*
  R banding, 185*b*
V-myc avian myelocytomatosis viral-related oncogene, neuroblastoma-derived, 381
VNTR. *See* Variable-number tandem repeat (VNTR)
Von Hippel–Lindau gene, *VHL* (3p26), 380–381
V-Ros Avian UR2 Sarcoma Virus Oncogene Homolog 1 (*ROS1*) Proto-Oncogene (6q22.1), 381
VZV. *See* Herpes virus (VZV)

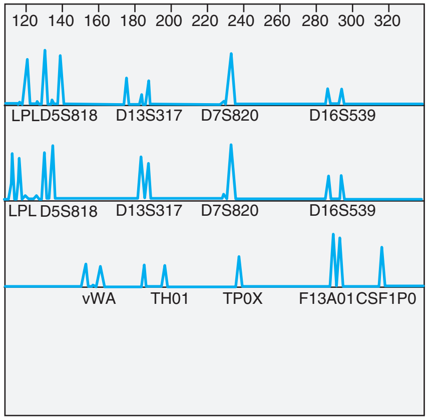
W
Waddington, Conrad, 65
Waldenström macroglobinemia, 406*t*
Warburg-Christian method, 93*b*
Watson, James, 3, 44*b*
Wells, 106
Western blot, 122–123, 123*b*, 127*b*, 127*f*, 133*f*
Whole chromosome paints, 193, 194*f*
Whole-genome amplification, 165, 165*f*, 166, 166*f*
Wollman, Elie, 23

X
Xenogeneic transplant, 284*b*
X-linked, 346
X-linked genes, 348
X-linked recessive diseases, 348*f*
xMHC. *See* Extended MHC locus (xMHC)

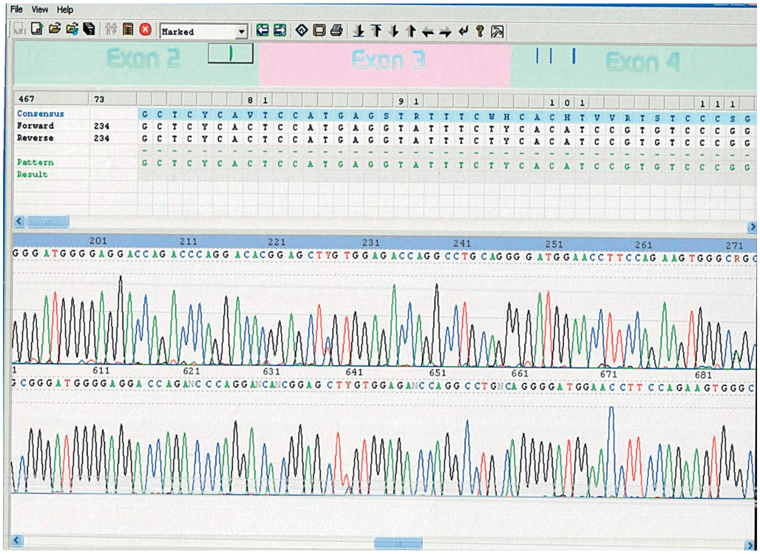
Y
Y-STR, 278–282
  autosomal STRs, 278
  genotypes, 281*t*
  locus information, 280*t*–281*t*
  matching with Y-STRs, 279, 281–282
    discriminatory capacity, 279
    haplotype diversity, 279
    surname test, 282
  paternal lineage test, 279
  purpose of, 268

Z
Zamecnik Paul, 46–47
  amino acid charging studies, 46
  ribosome studies, 47
Zimmerman, Steven B., 12*b*
Zinc finger motif, amino acid structure, 42*b*
Zwitterions, 38

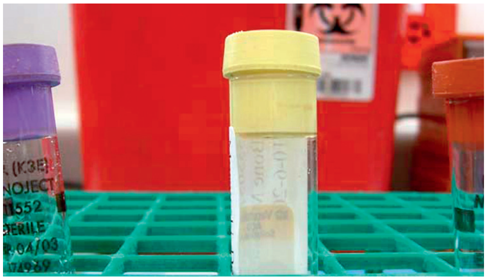
QUICK VISUAL REFERENCE FOR COMMON TECHNIQUES IN THE CLINICAL LABORATORY



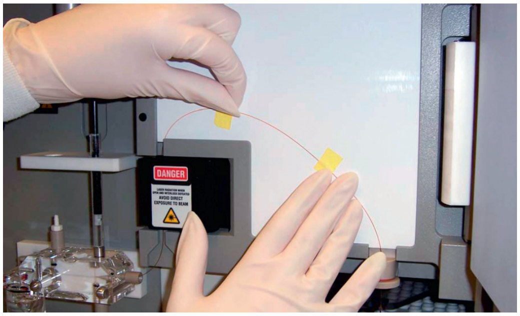
Screening of 16 loci for informative short tandem repeat alleles


File
View
Help
Marked
Exon 2
Exon 3
Exon 4
Consensus
Forward
Reverse
Pattern
Result

Sequence-based typing result for HLA-A gene



Ethylene diamine tetra acetic acid (EDTA) or acid citrate dextrose (ACD) tubes preferred for molecular analysis of blood or bone marrow specimens



Electrophoresis capillary replacement and instrument cleaning are examples of routine maintenance performed by laboratory personnel.

Continued from inside front cover